

8020866

CONTRIBUTIONS TO INFORMATION THEORY
WITH SPECIAL REFERENCE TO
SUBADDITIVITY AND OPTIMIZATION

by
VEMPATY NARASIMHAMURTHY

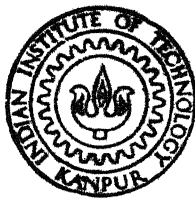
MATH

1987

D
NAR

CON

TH
MATH/1987/D
N 164c



DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
JULY, 1987

Doc. No. 106266

74
519.3
116640

MATH-1987-D-NAR-CON

To
my parents

CERTIFICATE

This is to certify that the matter embodied in the thesis entitled "Contributions to Information Theory with Special Reference to Subadditivity and Optimization" by Mr. V.N. Murthy for the award of the degree of Doctor of Philosophy of the Indian Institute of Technology, Kanpur is a record of bonafide research work carried out by him under joint supervision and guidance of Prof. J.N. Kapur and Dr. B.L. Bhatia. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

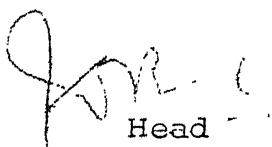
B.L. Bhatia 14.7.87
(J.N. Kapur) (B.L. Bhatia)
Prof. (Retd.) Asst. Prof.
Department of Mathematics
I.I.T., Kanpur


Course Certificate

This is to certify that Vempaty Narasimha Murthy has credited the following courses as the partial requirement for the Ph.D. programme in Mathematics

- M523 Measure Theory
- M602 Graduate Mathematics II
- (i) Functional Analysis
- (ii) Topology
- (iii) Measure Theory
- (iv) Linear Algebra
- M622 Holomorphy, Operator Theory and Two Norm Structure
- M601 Graduate Mathematics I
- (i) Complex Analysis
- (ii) Differential and Integral Equations
- M726 Univalent and Multivalent Functions
- M822 Topics in Analysis.

Vempaty Narasimha Murthy was admitted to the candidacy of the Ph.D. degree in April 1982 after completing his qualifying examinations.


Head
Department of Mathematics
I.I.T., Kanpur


Convenor
Departmental Postgraduate
Committee

Acknowledgements

I was very fortunate to have the opportunity to do my dissertation work under the supervision of such eminent and renowned mathematicians as Prof. J.N. Kapur and Dr. B.L. Bhatia. I can never thank enough or be able to express my gratitude to my supervisors in words for all the stimulating discussions we had, or further motivation they imparted to me. Mrs. Kapur has been very kind to me and I will remember the association throughout my life.

I take this opportunity to express my gratitude to my friends Dr. T.V. Prabhakar, Dr. H.C. Karnick, S. Shri V. Arvind, A. Subba Rao for their affectionate understanding of my troubles and for their endeavours to shed light on some dark patches in my thinking. I can never forget my dears, Dr. C. Venkobachar, Mrs. Sudha Venkobachar who literally pushed me towards my goal. Of course, my little friend Sandhya Jyothi has always been a delightful interlude.

My parents and brothers have all been the inspiration that I needed at times. My co-workers Mrs. M.S. Dhande and Mr. G.P. Tripathi have always extended the helping hand that I required.

Thanks are due to Mr. G.L. Mishra and Mr. S.K. Tewari for their excellent typing of the manuscript. Lastly but not in the least I thank whole heartedly all those countless well wishers of mine both among the staff and the faculty of I.I.T., Kanpur

July - 1987

V. N. MURTHY

CONTENTS

	Title	Page
CHAPTER 1 :	INTRODUCTION	1
1.1 :	Entropy	1
1.2 :	Directed Divergence, Independence Inequality and Optimization	3
1.3 :	Inaccuracy	7
1.4 :	Subadditivity, Superadditivity and Measures of Dependence	8
CHAPTER 2 :	A COMPARATIVE STUDY OF VARIOUS MEASURES OF INFORMATION	11
2.1 :	Measures of Entropy	11
2.2 :	Measures of Directed Divergence	51
2.3 :	Measures of Inaccuracy	78
CHAPTER 3 :	INDEPENDENCE INEQUALITY AND SUB- ADDITIVITY FOR MEASURES OF DIRECTED DIVERGENCE	114
3.1 :	Introduction	114
3.2 :	The I.I. for Measures of Directed Divergence	115
CHAPTER 4 :	SUBADDITIVITY, SUPERADDITIVITY AND MEASURES OF DEPENDENCE	131
4.1 :	Introduction	131
4.2 :	On the Relationship Between Super- additivity and Subadditivity of Renyi's and Havrda-Charvats' Measures of Entropy	136
4.3 :	Subadditivity and Superadditivity of Renyi's Measure of Entropy	139
4.4 :	Subadditivity and Superadditivity of Havrda-Charvats' Measure of Entropy	149
4.5 :	Subadditivity and Superadditivity of Kapur Aczel-Daroczy Measure of Entropy	152

CHAPTER 5 :	TWO OPTIMIZATION PROBLEMS IN INFORMATION THEORY	156
5.1 :	Geometry of the Probability Space	157
5.2 :	First Optimization Problem	160
5.3 :	An Alternative Proof of the Results Regarding the Maximum and Minimum Values of Θ	162
5.4 :	Second Optimization Problem	165
5.5 :	First Problem using Generalized Measures of Directed Divergence	166
5.6 :	Second Problem using Generalized Measures of Directed Divergence	170
5.7 :	Geometric Interpretation	174
CHAPTER 6 :	SOME MORE OPTIMIZATION PROBLEMS	175
6.1 :	Problem 1	176
6.2 :	Problem 2	179
6.3 :	Problem 3	181
6.4 :	Problem 4	183
6.5 :	Problem 5	186
6.6 :	Problem 6	190
6.7 :	Problem 7	192
REFERENCES		194

Synopsis

Contributions to Information Theory with Special Reference to Subadditivity and Optimization

A Thesis Submitted

In Partial Fulfilment of the Requirements

for the degree of

Doctor of Philosophy

By

VEMPATY NARASIMHA MURTHY

to the

Department of Mathematics

Indian Institute of Technology, Kanpur

In 1948, Shannon had introduced a mathematical model of communication system and a function to quantify the information given by a random variable. His measure of information or entropy is given by

$$H(P) = - \sum_{i=1}^n p_i \ln p_i$$

where $P = (p_1, p_2, \dots, p_n)$, $0 \leq p_i \leq 1$, $\sum_{i=1}^n p_i = 1$ is the probability distribution of a discrete random variable X . $H(P)$ is built based on some intuitively necessary axioms of the concept of information. In addition to these axioms, $H(P)$ satisfies properties like subadditivity, concavity with respect to P , recursivity property and so on.

Any measure of information is expected to satisfy most of these properties, if not all.

Renyi, 1961, Havrda-Charvat, 1967, Behara-Chawla, 1974, Aczel and Daroczy, 1963, Kapur, 1967, 84,85,86, Van der Lubbe et al, 1984, Sharma and Taneja, 1975, and Sharma and Mittal, 1975 have all proposed new and generalized measures of entropy with one or more parameters. In Chapter 2 section 1, we investigate which of these measures satisfy what properties and also give proofs in the cases where no attempts have been made so far.

In section 2, we deal with measures of directed divergence and in section 3, we deal with measures of inaccuracy. In both the cases, the idea is the same: to find out exhaustively the properties of the measure in consideration. Several proofs of properties and counter-proofs of properties by examples have been given. At the end of each section, a comprehensive table is provided with measures vs properties, which is helpful in finding out at a glance if a particular measure satisfies a particular property or not.

In 1977, A.B. El-Sayeed had introduced an inequality called the Independence Inequality for measures of entropy. It reads,

$$H(P*Q) \leq H(PQ)$$

or in words, the entropy of a joint probability scheme is maximum for the product distribution of the marginal

probability distributions. In Chapter 3 we introduce the Independence inequality for directed divergence measures and obtain various results for i) Renyi's measure of directed divergence, ii) Havrda-Charvats' measure of directed divergence iii) Kapur's measure of directed divergence of order α and type β and iv) Sharma and Guptas' measure of directed divergence. We also establish the connection between subadditivity and independence inequality for directed divergence measures.

In Chapter 4, we consider the properties of subadditivity and superadditivity for Havrda-Charvat and Renyi's measures of entropy. Here we consider these measures as families of measures, rather than single measures. Hence we are able to clearly discuss the subadditivity and superadditivity for some values of the parameters, in both the cases. We also categorise the set of probability distributions into two classes. For type I distributions, we establish a point α^* where Renyi's measures of entropy change from being subadditive to superadditive. We also show the usage of this critical point α^* by building a measure of dependence between probability distributions. We do the same for Havrda-Charvat's measure of entropy also.

In Chapters 5 and 6 we consider some applications of the concept of directed divergence in Information Theory. The classical distance between two points, x, y in Euclidean

space is a function, $d(.,.)$ satisfying

- i) $d(x,y) \geq 0 \quad \forall x \text{ and } y$
- ii) $d(x,y) = 0$ if and only if $x = y$
- iii) $d(x,y) = d(y,x) \quad \forall x \text{ and } z$ (symmetry property)
- iv) $d(x,z) \leq d(x,y) + d(y,z) \quad \forall x,y \text{ and } z$ (Triangular inequality).

Now we can consider $D(P;Q)$, the directed divergence between two probability distributions P and Q , as the distance between two points P and Q in the space of probability distributions. By definition $D(P;Q)$ satisfies i) and ii) above, but iii) and iv) are relaxed. Then with $D(P;Q)$ as the distance and probability distributions as points we can have a geometry of probability space.

In Chapter 5 we consider two optimization problems. The first problem is motivated by an optimization problem solved by Kullback [S. Kullback, 'Information Theory and Statistics', John Wiley, New York, 1959] in which he minimized $D(P:R)$ subject to $D(P:R) - D(P:Q) = \theta$ where θ is a fixed constant, and $D(P:R) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i}$, the Kullback-Libler measure of directed divergence. He discussed the solution for θ lying between 0 and 1, but it is actually valid for a larger range of values of θ . We find this range precisely. We consider the optimization problem in which we find the extremum values of θ .

The second optimization problem is one in which we find the maximum and minimum value of

$$\varphi = \sum_{j=1}^m \lambda_j D(P:Q_j)$$

where $\lambda_j > 0 \forall j$ and $\sum_j \lambda_j = 1$, Q_1, Q_2, \dots, Q_m are m given probability distributions. Later we solve both the optimization problems using more generalized measures of directed divergence as distance functions.

In Chapter 6 we consider seven optimization problems in probability space, which have very interesting Euclidean Geometric equivalents. These problems are solved by J.N. Kapur, in a recent paper, using Kullback-Leibler's measure of directed divergence. We generalize the solutions of these problems by using

$$D^\alpha(P, Q) = \frac{1}{\alpha-1} \left[\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} - 1 \right]$$

the Havrda-Charvat measure of directed divergence. In cases where closed form solutions are very complicated, we study some interesting particular cases with the help of numerical computations.

Chapter 1

Introduction

1.1 Entropy

In 1948, C.E. Shannon had constructed a mathematical theory for communication systems. The rudiments of a communication system are i) the source from which messages originate ii) the channel through messages are transmitted and iii) the receiver which receives the transmitted messages.

The source and the receiver had been treated as random experiments with either continuous or discrete outcomes. The amount of uncertainty, which Shannon called entropy or information about the outcome of a random experiment played a key role in Shannon's model for communication systems.

Let X be a discrete random variable with $P = (p_1, p_2, \dots, p_n)$ where $0 \leq p_i \leq 1$, $i = 1, 2, \dots, n$ and $\sum_{i=1}^n p_i = 1$ be its probability distribution. Then the following four axioms were used by Shannon to build his measure of entropy, $H(P)$:

I. $H(1/n, 1/n, \dots, 1/n) = f(n)$ is a monotonically increasing function of n ($n = 1, 2, \dots$).

II. $f(n \cdot m) = f(n) + f(m)$, ($n, m = 1, 2, \dots$).

III. $H(p_1, p_2, \dots, p_n) = H(p_1 + \dots + p_n, p_{n+1} + \dots + p_n)$

$$\begin{aligned}
& + (p_1 + \dots + p_n) H\left(\frac{p_1}{\sum_{i=1}^n p_i}, \dots, \frac{p_r}{\sum_{i=1}^n p_i}\right) \\
& + (p_{r+1} + \dots + p_n) H\left(\frac{p_{n+1}}{\sum_{i=r+1}^n p_i}, \dots, \frac{p_n}{\sum_{i=r+1}^n p_i}\right)
\end{aligned}$$

(III is called 'grouping axiom').

IV. $H(p, 1-p)$ is a continuous function of p .

Shannon obtained

$$H(P) = - \sum_{i=1}^n p_i \log p_i \quad (1.1)$$

as the measure of entropy satisfying the four axioms. In 1957 Kinchin proved that any function satisfying I, II, III and IV above must be a constant multiple of Shannon's measure of entropy.

In the three ensuing decades after Shannon's discoveries, various scientists and engineers had obtained many more measures of entropy by generalizing some of Shannon's axioms or by deleting some axioms altogether. But not all the measures proposed later satisfied the properties satisfied by Shannon's measure.

We discussed in Chapter 2, section 1, the properties possessed by Shannon's measure of entropy apart from the four mentioned before. We also tabulated some of the more prominent measures of entropy and studied their properties. That is, we

checked for the generalized measures the properties possessed by Shannon's measure. The idea was not to contradict anybody (because many authors didn't publish results regarding the validity of these properties for their respective measures) but only to highlight positive and negative aspects of these measures of entropy. After all, there is no good in treating a measure, which can also assume negative values, as a measure of entropy.

1.2 Directed Divergence, Independence Inequality and Optimization

S. Kullback and R.A. Leibler had introduced in 1951 a concept which they called mutual information. If $P=(p_1, \dots, p_n)$ and $Q = (q_1, \dots, q_n)$ were two discrete probability distributions, then Kullback-Leiblers' measure of mutual information was given by

$$D(P:Q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} \quad (1.2)$$

where $q_i \neq 0$ if $p_i \neq 0$, $i = 1, 2, \dots, n$ and $0 \cdot \log 0$ was treated as zero.

$D(P:Q)$ satisfied the following properties :

- (i) $D(P:Q) \geq 0$ for all P and Q
- (ii) $D(P:Q) = 0$ iff $P = Q$
- (iii) $D(P:Q)$ is a convex function of P and of Q .

$D(P:Q)$ can be compared to the classical distance function $d(x,y)$ where x and y are points in any Euclidean space.

That is in a way we could treat $D(P:Q)$ as the distance (or the directed divergence) of P from Q . Here the symmetry property and triangular inequality of $d(x,y)$ were relaxed. So $D(P:Q)$ might not be equal to $D(Q:P)$. In fact that was the reason why $D(P:Q)$ was referred to as the directed divergence of P from Q .

There were very many directed divergence measures in the literature of information theory. But not all of them satisfied all the properties that a directed divergence measure was expected to satisfy and indeed there were measures which violated even the non-negativity criteria. The property of convexity of the directed divergence measures was insisted upon because then the local minimum of $D(P:Q)$ over P would be the global minimum. This property would come very handy in solving optimization problems in probability space which arise in many fields of research.

In Chapter 2, section 2 we tabulated the properties which were essential for a directed divergence and checked them for measures of directed divergence, which were also tabulated there.

In Chapter 3 we proposed an inequality, the independence inequality for directed divergence measures. A.B.El-Sayeed had introduced the independence inequality

$$H(P * Q) \leq H(PQ) \quad (1.3)$$

for measures of entropy. The meaning of (1.3) in words was that the entropy of a joint probability scheme was maximum for the product distribution of the marginal distributions.

The independence inequality for directed divergence measures had been defined by us as

$$D(P * Q : R * S) \leq D(PQ : RS) \quad (1.4)$$

where $P * Q$ and $R * S$ were joint probability distributions with P, Q and R, S being the marginal distributions. The meaning of (1.4) in words was that the directed divergence of a joint probability distribution from another joint probability distribution was maximum when the products of their marginal probability distributions were considered.

We considered four measures of directed divergence : Renyi's measure, Havrda-Charvats', Kapur's measure of order α and type β and Sharma and Tanejas' measures of directed divergence. We first established that either the Renyi and the Havrda-Charvats' measures of directed divergence either both satisfied the independence inequality or neither satisfied it. Then we provided examples of joint probability distributions for which the independence inequality was not satisfied by any of the four measures. So, we had established the fact that the independence inequality was as much a property of the distributions as it was of a directed divergence measure. Then we constructed some general probability distributions for which all the four measures satisfied the independence inequality.

The connection between the independence inequality and subadditivity property of the directed divergence measures was also established. If a directed divergence measure satisfied

the independence inequality and also was additive, then it satisfied the subadditivity law also.

In Chapter 4 we considered two optimization problems in probability space. In the first problem we obtained the maximum and minimum of

$$\theta = D(P:Q) - D(P:R) \quad (1.5)$$

where Q and R were given distributions, $D(P:Q)$ was i) Kullback-Leibler, ii) Havrda-Charvat, iii) Kapur, iv) Ferrari measure of directed divergence Kullback maximized $D(P:Q)$ s. to

$$\theta = D(P:Q) - D(P:R) \quad (1.6)$$

where θ was a real constant, $0 \leq \theta \leq 1$. But we had shown that his solution is valid for a bigger range of values of θ . Infact we had derived motivation for problem 1 from Kullback's problem.

In Problem 2 we find maximum and minimum of

$$\varphi = \sum_{j=1}^m \lambda_j D(P:Q_j) \quad (1.7)$$

where Q_j , $j = 1, 2, \dots, m$ were given probability distributions and $0 < \lambda_j < 1$, $\sum_{j=1}^m \lambda_j = 1$ were constants. Here $D(P, Q_j)$ was i) Kullback-Leiber, ii) Havrda-Charvat and iii) Ferrari measure of directed divergence.

We had also provided alternate proofs for Kullback's results.

In Chapter 6 we had considered seven more optimization problems with Havrda-Charvats' measure of directed divergence as the distance function. Kapur had earlier solved these problems using Kullback-Leiblers' measure of directed divergence as the distance function. We had also illustrated our results with some numerical computations.

1.3 Inaccuracy :

Let $P = (p_1, p_2, \dots, p_n)$ be the correct probability distribution of a stochastical experiment. Let $Q = (q_1, q_2, \dots, q_n)$ be the probability distribution which was asserted to be the right distribution for the experiment. It was really interesting to know how much inaccurate the assertion was. D.F. Kerridge in 1961, characterized a measure for inaccuracy $I(P:Q)$. He had based his measure on the following four axioms :

- I. I is continuous in both p_i and q_i for all $i = 1, 2, \dots, n$.
- II. When n equally likely outcomes are stated to be equally likely then I is a monotonic increasing function of n .
- III. If a statement is broken down in to a number of subsidiary statements the inaccuracy of the original statement is a weighted sum of the inaccuracies of the subsidiary statements.
- IV. The inaccuracy of a statement is unchanged if two alternatives about which the same assertion is made are combined.

Kerridge had established that all the four axioms are

satisfied if and only if

$$I(P:Q) = -k \sum_{i=1}^n p_i \log q_i \quad (1.8)$$

where k is a multiplicative constant.

Again there was a plethora of inaccuracy measures in the literature. Kerridge's measure satisfied several useful properties apart from the four axioms mentioned earlier. However some of the measures obtained by other authors did not satisfy all of these properties. For example III, the recursivity (or the grouping) axiom is satisfied only by Kerridge's measure. There were some measures which also assumed negative values, but there was no plausible interpretation for negative inaccuracy. We had listed the more prominent measures of inaccuracy and also the properties which were desirable for an inaccuracy measure, in section 3, Chapter 2. There we had endeavoured to verify the properties for each measure listed. In the cases where a measure did not satisfy a particular property, a counter-example had been given.

1.4 Subadditivity, Superadditivity and Measures of Dependence :

Shannon's measure of entropy satisfied the subadditivity law,

$$H(P * Q) \leq H(P) + H(Q) \quad (1.9)$$

and Renyi's measure of entropy of order α satisfied the super-additivity law,

$$H_{\alpha}(P * Q) \geq H_{\alpha}(P) + H_{\alpha}(Q) \quad (1.10)$$

for some values of α and some distribution $P * Q$.

Renyi's measure of entropy or any other measure of entropy of order α and type β was not a single measure. Infact it was a family of measures, one measure corresponding to a value of α or β . Though we knew that subadditivity property was not satisfied in general by Renyi's (or Havrda-Charvats' measure or any other measure with one or more parameters), it was really interesting to know for what range of values of α , and for what type of distributions it satisfied the subadditivity law.

In Chapter 4, we categorized the set of joint probability distributions into two classes, Type I and Type II. We established that for Type I distributions, there existed a point α_* s.t. $1 \leq \alpha_*$ and H_{α} was subadditive for all $\alpha \geq \alpha_*$ and superadditive for $0 \leq \alpha < \alpha_*$. However we were unable to rule out the possible existence of more than one α_* , but we proved that if there existed more than one α_* , then there was an even number of such numbers.

Now from the relation

$$H_{\alpha}(P * Q) = H_{\alpha}(P) + H_{\alpha}(Q) - (1-\alpha)H_{\alpha}(P)H_{\alpha}(Q) \quad (1.11)$$

we got that $\alpha_* = 1$ if $P * Q = PQ$ because H_{α} is additive. That is P and Q are independent, $\alpha_* = 1$. From the several calculations we had made we observed that α_* was closer to unity when P and Q

were close to being independent and α_* was significantly farther from unity if P and Q were far from being independent. A natural consequence of our work was to use α_* to measure the dependence between P and Q via $P * Q$. We had proposed to do it by evaluating $(\alpha_* - 1)$.

Similarly we had established another measure of dependence, $(1 - \alpha^*)$ by using the subadditivity and superadditivity properties of Havrda-Charvats' measure of entropy. Here we noted that irrespective of the type of the distribution, there existed at least one $\alpha^* \leq 1$ such that for $0 \leq \alpha \leq \alpha^*$, $H^\alpha(P * Q)$ is subadditive, while for $\alpha \geq \alpha^*$, $H^\alpha(P * Q)$ is superadditive.

Chapter 2

A Comparative Study of Various Measures of Information

Introduction : In this chapter we study various measures of entropy, directed divergence and inaccuracy, with respect to their properties. We take up these concepts in a section each. At the beginning of each section a list of various measures of the respective concept is provided. Then a list of properties an ideal measure of that concept is expected to satisfy. Finally a table with measures against their properties is compiled. This helps in finding out which measure satisfies what properties, at a glance.

We only provide proofs or counter-examples for those results which haven't been proved or counter proved. The proofs of other results are omitted.

Section 1 deals with the measures of entropy, section 2 with the measures of directed divergence and section 3 with the measures of inaccuracy.

2.1 Measures of Entropy

2.1.1 List of Measures of entropy :

The following measures of entropy are considered in this section :

Let $P = (p_1, p_2, \dots, p_n)$ be a discrete probability distribution, $p_i \geq 0 \forall i$ and $\sum_{i=1}^n p_i = 1$,

1) Shannon's [56] :

$$H(P) = - \sum_{i=1}^n p_i \ln p_i$$

2) Renyi's [65] measure of order α

$$H_{\alpha}(P) = \frac{1}{1-\alpha} \ln \sum_{i=1}^n p_i^{\alpha} \quad \alpha \neq 1$$

3) Havrda and Charvats' [8] measure of degree α

$$H^{\alpha}(P) = \frac{1}{1-\alpha} \left\{ \sum_{i=1}^n p_i^{\alpha} - 1 \right\} \quad \alpha \neq 1$$

4) Kapur's [9] measure of order α and type β

$$H_{\alpha, \beta}(P) = \frac{1}{1-\alpha} \ln \left\{ \frac{\sum_{i=1}^n p_i^{\alpha+\beta-1}}{\sum_{i=1}^n p_i^{\beta}} \right\} \quad \alpha \neq 1.$$

This measure is independently obtained by Aczel and Daroczy [1]. Therefore, it will be referred to as Kapur-Aczel and Daroczy measure of information of order α and type β .

5) Kapur's [17] four families of measures of entropy :

$$(i) \quad H_a(P) = - \sum_{i=1}^n p_i \ln p_i + \frac{1}{a} \sum_{i=1}^n \{ (1+ap_i) \ln(1+ap_i) - ap_i \}, a > 0$$

$$(ii) \quad H_b(P) = - \sum_{i=1}^n p_i \ln p_i + \frac{1}{b} \left\{ \sum_{i=1}^n (1+bp_i) \ln(1+bp_i) - (1+b) \ln(1+b) \right\} \quad b \geq 0$$

$$(iii) H_c(P) = - \sum_{i=1}^n p_i \ln p_i + \frac{1}{c^2} \left\{ \sum_{i=1}^n (1+cp_i) \ln(1+cp_i) - c \right\}$$

$$(iv) H_k(P) = - \sum_{i=1}^n p_i \ln p_i + \frac{1}{k^2} \left\{ \sum_{i=1}^n (1+kp_i) \ln(1+kp_i) - (1+k) \ln(1+k) \right\}$$

6) Behara and Chawlas' [3] γ -entropy :

$$H_{\gamma}(P) = \frac{1 - \left(\sum_{i=1}^n p_i^{1/\gamma} \right)^{\gamma}}{1 - e^{1-\gamma}} \quad \gamma \neq 1, \gamma > 0$$

7) Sharma and Tanejas' [57] two parametric measures of entropy:

(i) measure of order α and type β :

$$H^{\alpha, \beta}(P) = \frac{1}{(2^{1-\alpha} - 2^{1-\beta})} \left\{ \sum_{i=1}^n p_i^{\alpha} - \sum_{i=1}^n p_i^{\beta} \right\} \quad \alpha \neq \beta$$

(ii) logarithm measure : $H_L(P) = -2^{\alpha-1} \sum_{i=1}^n p_i^{\alpha} \ln p_i$

(iii) Sine measure : $H_S(P) = \frac{2^{\alpha-1}}{\sin \beta} \sum_{i=1}^n p_i^{\alpha} \sin(\beta \log p_i)$

8) Rathie's [51] measure of entropy with $(m+1)$ parameters :

$$H_{\alpha}^{\beta_1, \dots, \beta_n}(P) = \frac{1}{2^{1-\alpha} - 1} \left\{ \sum_{i=1}^n \frac{p_i^{\alpha+\beta_i-1}}{p_i^{\beta_i}} - 1 \right\} \quad \alpha \neq 1.$$

9) Sharma and Mittals' [58] measure of entropy :

$$H_{a,b}(P) = \frac{1}{(1-a)^b} \left\{ \left(\sum_{i=1}^n p_i^a \right)^b - 1 \right\}, \quad a > 0, a \neq 1.$$

10) J.C.A. Van der Lubbe et als' [61] generalized measures :

$$(1) H_n^1(P; \rho, \sigma, \delta) = - \delta \log \left\{ \sum_{i=1}^n p_i^{\rho} \right\}^{\sigma}$$

$$(ii) H_n^2(P; \rho, \sigma, \delta) = \delta \{1 - (\sum_{i=1}^n p_i^\rho)^\sigma\}$$

$$(iii) H_n^3(P; \rho, \sigma, \delta) = \delta \{(\sum_{i=1}^n p_i^\rho)^{-\sigma} - 1\}$$

where $(\rho, \sigma) \in D = \{(\rho, \sigma) | \rho > 1, \sigma > 0 \vee 0 < \rho < 1, \sigma < 0\}$.

11) Kapur's [40] generalized measure of entropy :

$$H_{\alpha, \beta}^b(P) = \frac{(\sum_{i=1}^n p_i^\alpha)^b - (\sum_{i=1}^n p_i^\beta)^b}{(\beta - \alpha)b} \quad \alpha \neq \beta, b \neq 0, \alpha, \beta > 0.$$

12) Kapur's [37] three parametric measures of Entropy :

$$(i) H_k^1(P) = \frac{1}{\alpha - \beta} \{(\frac{\lambda}{n} + (1-\lambda))^{1-\alpha} - (\frac{\lambda}{n} + (1-\lambda))^{1-\beta}\} \\ - \frac{1}{\alpha - \beta} \{ \sum_{i=1}^n p_i^\alpha (\frac{\lambda}{n} + (1-\lambda)p_i)^{1-\alpha} + \sum_{i=1}^n p_i^\beta (\frac{\lambda}{n} + (1-\lambda)p_i)^{1-\beta} \}$$

$$(ii) H_k^2(P) = \frac{1}{(\alpha-1)a} \{(\frac{\lambda}{n} + (1-\lambda))^{(1-\alpha)a} - \sum_{i=1}^n p_i^\alpha (\frac{\lambda}{n} + (1-\lambda)p_i)^{(1-\alpha)a}\}$$

$$(iii) H_k^3(P) = \frac{1}{(\alpha-1)} \ln \frac{\{(\frac{\lambda}{n} + (1-\lambda))^{1-\alpha}\}}{\sum_{i=1}^n p_i^\alpha (\frac{\lambda}{n} + (1-\lambda)p_i)^{1-\alpha}}.$$

That completes the list of measures of entropy. We now provide the list of properties of measures of entropy in the following subsection.

2.1.2 List of Properties of Measures of Entropy :

For the justification of these properties, we refer one to [13,22,24,27,33,40]. All these papers contain discussions on the properties of measures of entropy. The following properties are verified for the measures of entropy listed in 1.1. Let

$A_n = \{(p_1, \dots, p_n) \mid 0 \leq p_i \leq 1, \sum_{i=1}^n p_i = 1\}$ be the set of all discrete probability distributions with n outcomes and let $H_n(P)$ be any measure of entropy for $P \in A_n$.

- 1) $H_n(P)$ is a continuous function of the probabilities p_1, \dots, p_n .
- 2) $H_n(P)$ is a symmetric function of the probabilities.
- 3) $H_n(P)$ is zero for degenerate $P \in A_n$ and this is the minimum value attained by $H_n(P)$ over A_n .
- 4) $H_n(P)$ is maximum when all the probabilities are equal. That is when $P = U \in A_n$.
- 5) $\max_{P \in A_n} H_n(P)$ is an increasing function of n .
- 6) If an outcome of zero probability is added to the experiment, the amount of entropy remains the same. This property is called expansibility property.
- 7) If $P \in A_n$ and $Q \in A_m$ are independent and PQ is their product distribution in A_{mn} then we have $H_{mn}(PQ) = H_n(P) + H_m(Q)$. This property is referred to as additivity.
- 8) For any $P \in A_n$ and $Q \in A_m$ and $P*Q$ their joint distribution in A_{mn} then we have $H_{mn}(P*Q) \leq H_n(P) + H_m(Q)$. This property is referred to as subadditivity.
- 9) $H_n(P)$ is a concave function of P . This property of an information measure is desired because it ensures that the local minimum of it is the global minimum. Also when a function is concave, Lagrange's multipliers method always yields minimum.

- 10) If any parameters are included in the measure then it is a monotonic function with respect to them.
- 11) It is also either concave or convex with respect to the parameters.

We now go to the study of the measures of entropy.

Analysis

2.1.3 Shannon's Measure : It is given by $H(P) = - \sum_{i=1}^n p_i \ln p_i$.

It satisfies all the properties (1) to (9). Because there is no parameter involved in this measure, there is no question of verifying properties (10) and (11). Its maximum value is $\ln(n)$ and it is a concave function of P . It is additive and satisfies the subadditivity property. In addition to the above, Shannon's measure satisfies the recursivity property,

$$H_n(p_1, p_2, \dots, p_n) = H_{n-1}(p_1+p_2, p_3, \dots, p_n) + (p_1+p_2) H_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right).$$

2.1.4 Renyi's Measure of Entropy : It is defined as

$$H_\alpha(P) = \frac{1}{1-\alpha} \ln \left\{ \sum_{i=1}^n p_i^\alpha \right\} \quad \alpha \neq 1.$$

It is easily seen that $H_\alpha(P) \rightarrow H(P)$ as α approaches unity. Obviously it is a continuous symmetric function of the probabilities. It has maximum value equal to $\ln(n)$ which is attained for the uniform distribution $U = (\frac{1}{n}, \dots, \frac{1}{n})$. It's minimum value n -times is zero and is attained for any degenerate distribution. Renyi's

measure of entropy is additive but it is not subadditive. We deal with this property in complete detail in Chap. 4. But here we give an example of A.B.El-Sayeed [5] to justify our claim.

Example 2.1.4.1 : Let $\alpha = 2$, $n = 2$, $m = 2$

$$(\pi_{ik}) = \begin{pmatrix} 0.02 & 0.08 \\ 0.27 & 0.63 \end{pmatrix} \quad P = (0.1, 0.9) \text{ and } Q = (0.29, 0.71) .$$

We then have $H_\alpha(\pi) = 1.4822$ and $H_\alpha(P) + H_\alpha(Q) = 1.4583$. Thereby we get that $H_\alpha(\pi_{ik}) > H_\alpha(P) + H_\alpha(Q)$. Thus for $\alpha = 2$, $H_\alpha(P)$ is not subadditive.

Renyi's measure of entropy satisfies the expansibility property. It's concavity property with respect to P is discussed in B. Bessat and A. Raviv [4]. Their results state that (i) $H_\alpha(P)$ is strictly concave w.r.to P for $0 < \alpha \leq 1$. (ii) For $n = 2$, $H_\alpha(P)$ is strictly concave for $0 < \alpha \leq 2$. (iii) $H_\alpha(P)$ is pseudoconcave for values of $\alpha > 0$.

Kapur [34] has established that $H_\alpha(P)$ is a pseudoconvex function of α for $0 < \alpha < 1$ and pseudoconvex function of α for $\alpha > 1$. It is also established there that $H_\alpha(P)$ is monotonically decreasing function of α .

Renyi's measure does not have an easy recursive property. By the following relation between Renyi's and Havrda and Charvat's measures of entropy,

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \{(\alpha-1)H^\alpha(P) + 1\}$$

and a recursive property for the latter measure, we can deduce a complicated expression for recursivity of $H_\alpha(P)$. We state and prove the recursivity of degree α for Havrda Charvats' measure of entropy in the next subsection.

With that we conclude the discussion of Renyi's measure of entropy.

2.1.5 Havrda Charvats' Measure of Entropy : It is given by

$$H^\alpha(P) = \frac{1}{1-\alpha} \left\{ \sum_{i=1}^n p_i^\alpha - 1 \right\} \quad \alpha \neq 1.$$

We can easily see that $H^\alpha(P)$ tends to $H(P)$ as α approaches 1. $H^\alpha(P)$ is continuous, symmetric function of the probabilities. Its minimum value is zero and is attained for degenerate distribution. $H^\alpha(P)$ attains its maximum value for the uniform distribution and the max. value is $\frac{n^{1-\alpha}-1}{1-\alpha}$ which is not independent of α unlike the max. value of $H_\alpha(P)$. But again $\max_P H^\alpha(P)$ approaches $\ln(n)$ as $\alpha \rightarrow 1$ and it is an increasing function α .

Unlike Renyi's and Shannon's measures, Havrda and Charvat's measure of entropy is non-additive. For independent probability distributions the following relation holds for $H^\alpha(P)$:

$$H_{nm}^\alpha(PQ) = H_n^\alpha(P) + H_m^\alpha(Q) + (1-\alpha)H^\alpha(P)H^\alpha(Q).$$

Now it follows from the above relation that $H^\alpha(P)$ is additive for $\alpha = 1$. But for $\alpha \neq 1$ unless either of the distributions P and Q is degenerate $H^\alpha(P)$ is not additive. Therefore

$H^\alpha(P)$ is additive only for $\alpha = 1$.

Subadditivity is also not universally true for $H^\alpha(P)$. In fact we can deduce this fact from the above relation itself. If $\alpha < 1$ we can always find distributions which violate subadditivity rule. We discuss in Chap. 4 in complete detail the subadditivity property of $H^\alpha(P)$.

Like Shannon's measure and unlike Renyi's measure $H^\alpha(P)$ is concave for all values of α . We can easily verify this by considering the fact that $(\sum_{i=1}^n p_i^\alpha)$ is concave for $\alpha < 1$ and convex for $\alpha > 1$.

Now we shall consider recursivity for $H^\alpha(P)$.

Proposition 2.1.6.1 : $H^\alpha(P)$ satisfies the following relation for all values of α :

$$H_n^\alpha(p_1, p_2, \dots, p_n) = H_{n-1}^\alpha(p_1+p_2, p_3, \dots, p_n) + (p_1+p_2)^\alpha H_2^\alpha\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right).$$

(Note : This is not a new result but the proof is independently obtained by us).

Proof : Let $G_n(P) = G_n(p_1, \dots, p_n) = \sum_{i=1}^n p_i^\alpha$. Then we have

$$H_n^\alpha(P) = \frac{1}{1-\alpha} \{G_n(P) - 1\}.$$

We shall first obtain a recursive relation for $G_n(P)$. We have from the definition

$$G_{n-1}(p_1+p_2, p_3, \dots, p_n) = (p_1+p_2)^\alpha + \sum_{i=3}^n p_i^\alpha \text{ and}$$

$$G_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) = \frac{p_1^\alpha + p_2^\alpha}{(p_1+p_2)^\alpha}.$$

Now

$$(p_1+p_2)^\alpha G_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) + G_{n-1}(p_1+p_2, p_3, \dots, p_n) = G_n(p_1, \dots, p_n) \\ + (p_1+p_2)^\alpha.$$

Therefore

$$G_n(p_1, \dots, p_n) = G_{n-1}(p_1+p_2, p_3, \dots, p_n) + (p_1+p_2)^\alpha G_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) \\ - (p_1+p_2)^\alpha$$

and

$$H_n^\alpha(p_1, p_2, \dots, p_n) = \frac{1}{1-\alpha} \{G_n(p_1, \dots, p_n) - 1\} \\ = \frac{1}{1-\alpha} \{G_{n-1}(p_1+p_2, p_3, \dots, p_n) \\ - 1 + (p_1+p_2)^\alpha G_2\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right) - 1\} \\ = H_{n-1}^\alpha(p_1+p_2, p_3, \dots, p_n) + (p_1+p_2)^\alpha H_2^\alpha\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right).$$

That completes the proof.

Kapur [34] has established that $H^\alpha(p)$ is a monotonically decreasing function of α . He has also established that it is a pseudo-convex function of α for $0 < \alpha < 1$ and pseudo-concave for $\alpha > 1$. But we have obtained the following better result.

Proposition 2.1.5.2 : $H^\alpha(p)$ is a strictly convex function w.r.to

α for $0 < \alpha < 1$. We can not say anything about concavity or convexity of $H^\alpha(P)$ if $\alpha > 1$, better than Kapur's result.

Proof : Let $f(\alpha) = \frac{1}{1-\alpha} \{ \sum_{i=1}^n p_i^\alpha - 1 \}$ then we have

$$f'(\alpha) = \frac{\sum_{i=1}^n p_i^\alpha \log p_i}{(1-\alpha)} + \frac{\sum_{i=1}^n p_i^\alpha - 1}{(1-\alpha)^2} \text{ and}$$

$$f''(\alpha) = \frac{\sum_{i=1}^n p_i^\alpha (\log p_i)^2}{(1-\alpha)} + 2 \frac{\sum_{i=1}^n p_i^\alpha \log p_i}{(1-\alpha)^2} + 2 \frac{\sum_{i=1}^n p_i^\alpha - 1}{(1-\alpha)^3}.$$

Note that the term on the R.H.S. of the above expression is always positive and the second term negative irrespective of the value of α . But the first term is positive or negative depending on whether α is less or greater than 1. Now consider the case $0 < \alpha < 1$ and rewrite the above expression as

$$f''(\alpha) = \frac{1}{(1-\alpha)^2} \sum_{i=1}^n p_i^\alpha \{ (1-\alpha)(\log p_i)^2 + 2 \log p_i \} + 2 \frac{\sum_{i=1}^n p_i^\alpha - 1}{(1-\alpha)^3}.$$

Now we shall show that the first term in the last expression is positive. For let it be negative. Then $(1-\alpha)(\log p_i)^2 < 2 \log p_i$ for atleast one value of i . Let that value of i be denoted by k . That is

$$(1-\alpha)(\log p_k)^2 < 2 \log p_k$$

$$\text{or } \left(\frac{1-\alpha}{2}\right) \log p_k > 1 \quad \{\text{as } \log p_i < 0\}$$

$$\text{or } p_k > e^{\frac{2}{1-\alpha}} \quad \{\text{if natural logarithms were used}\}$$

But since we started with $0 < \alpha < 1$, which means $e^{\frac{2}{1-\alpha}} > 1$. This is a contradiction to the fact that $0 \leq p_i \leq 1 \forall i = 1, \dots, n$.

Therefore our assumption that $\frac{1}{(1-\alpha)^2} \sum_{i=1}^n p_i^\alpha \{ (1-\alpha)(\log p_i)^2 + 2 \log p_i \}$ is negative is wrong. Therefore we get that for $0 < \alpha < 1$ $f''(\alpha) > 0$ which means that $f(\alpha)$ is convex for $0 < \alpha < 1$.

It is obvious that we can not extend similar arguments for the case of $\alpha > 1$.

With that we conclude our discussion of the Havrda Charvats' measure of entropy.

2.1.6 Kapur-Aczel Daroczys' Measure of entropy of order α and type β

It is given by $H_{\alpha,\beta}(P) = \frac{1}{1-\alpha} \ln \left\{ \frac{\sum_{i=1}^n p_i^{\alpha+\beta-1}}{\sum_{i=1}^n p_i^\beta} \right\}$. We can obtain Renyi's measure of entropy from $H_{\alpha,\beta}(P)$ by taking $\beta = 1$ and Shannon's measure by taking $\beta = 1$ and letting α approach 1.

$H_{\alpha,\beta}(P)$ is obviously a continuous, symmetric function of p_i 's. Its minimum value is zero which occurs for any degenerate distribution. For uniform distributions its value is $\ln(n)$. But this is not the maximum value of it for all values of α and β . For a detailed discussion refer [9,10,15]. The range of values of α and β for which $H_{\alpha,\beta}(U)$ is maximum is as follows :

- (i) $\alpha > 1, 0 \leq \beta \leq 1$ and $\alpha+\beta \geq 2$ and
- (ii) $\alpha < 1, \beta > 1, 1 < \alpha+\beta < 2$

because when α, β take values satisfying (i) and (ii) above $H_{\alpha, \beta}(P)$ is pseudo-concave.

Apart from Shannon's and Renyi's measures of entropy $H_{\alpha, \beta}$ is the only other measure that is additive.

Obviously this measure is not subadditive for all values of α and β . We shall discuss the subadditivity property of this measure in detail in Chap. 4.

Having considered all other properties of $H_{\alpha, \beta}$ now we shall move on to its properties with respect to the parameters.

Proposition 2.1.6.1 : $H_{\alpha, \beta}(P)$ is a monotonically decreasing function of β for all values of α . But we can not conclude monotonic behaviour of $H_{\alpha, \beta}(P)$ with respect to α .

Proof : Let $f(\beta) = \frac{1}{1-\alpha} \log \left\{ \frac{\sum_{i=1}^n p_i^{\alpha+\beta-1}}{\sum_{i=1}^n p_i^{\beta}} \right\}$

we then have

$$f'(\beta) = \frac{1}{1-\alpha} \left\{ \frac{\sum_{i=1}^n p_i^{\alpha+\beta-1} \log p_i}{\sum_{i=1}^n p_i^{\alpha+\beta-1}} - \frac{\sum_{i=1}^n p_i^{\beta} \log p_i}{\sum_{i=1}^n p_i^{\beta}} \right\}.$$

Case (i) $\alpha > 1$: i.e., $\alpha + \beta - 1 > \beta$ which implies $p_i^{\alpha+\beta-1} \leq p_i^{\beta} \forall i=1, 2, \dots, n$.

But we have $0 \leq p_i \leq 1 \Rightarrow \log p_i \leq 0 \forall i=1, 2, \dots, n$. Therefore

we now have

$$\sum_{i=1}^n p_i^{\alpha+\beta-1} \log p_i \geq \sum_{i=1}^n p_i^{\beta} \log p_i \text{ and } \sum_{i=1}^n p_i^{\alpha+\beta-1} \leq \sum_{i=1}^n p_i^{\beta}$$

which will immediately give us $f'(\beta) \leq 0$ because $(1-\alpha) < 0$.

Case (ii) $0 < \alpha < 1$: Presently we have $(1-\alpha) > 0$ and the rest of the inequalities in case (i) are reversed giving $f'(\beta) \leq 0$ again.

Now consider the following example to verify our claim regarding the monotonicity of $H_{\alpha,\beta}(P)$ as a function of α .

Let $P = (0.5, 0.6)$, $\beta = 2$, $\alpha_1 = 0.5$, $\alpha_2 = 1.5$, $\alpha_3 = 2$, $\alpha_4 = 20$, $\alpha_5 = 50$. We get $g(\alpha_1) = 0.8646137$, $g(\alpha_2) = 0.2723305$, $g(\alpha_3) = 0.5376906$, $g(\alpha_4) = 0.2302494$ and $g(\alpha_5) = 0.2251079$ where $g(\alpha) = H_{\alpha,\beta}(P)$. Therefore $g(\alpha)$ is not a monotonic function of α . That completes the proof of Proposition 2.1.6.1.

Convexity or concavity of this measure w.r.to any of the two parameters is very difficult to establish owing to the fact that the expressions involved are very complicated. We shall now close our discussion of this measure.

2.1.7 Kapur's Four families of Measures of Entropy

These measures are given by

$$(i) \quad H_a(P) = - \sum_{i=1}^n p_i \ln p_i + \frac{1}{a} \sum_{i=1}^n \{ (1+ap_i) \ln(1+ap_i) - ap_i \} \quad a > 0$$

$$(ii) \quad H_b(P) = - \sum_{i=1}^n p_i \ln p_i + \frac{1}{b^2} \left\{ \sum_{i=1}^n (1+bp_i) \ln(1+bp_i) - (1+b) \ln(1+b) \right\}$$

$$(iii) \quad H_c(P) = - \sum_{i=1}^n p_i \ln p_i + \frac{1}{c^2} \left\{ \sum_{i=1}^n (1+cp_i) \ln(1+cp_i) - c \right\}$$

$$(iv) \quad H_k(P) = - \sum_{i=1}^n p_i \ln p_i + \frac{1}{k^2} \left\{ \sum_{i=1}^n (1+kp_i) \ln(1+kp_i) - (1+k) \ln(1+k) \right\}.$$

One salient feature of these four families of measures of entropy is that they are constructed with a view to satisfy all the important properties of a measure of entropy. Refer to Kapur [17] for a detailed discussion on properties of the measures belonging to these four classes.

However, we wish to point out an observation regarding the minimum value of any of the measures of these families. For degenerate distributions they attain their minimum. But unlike any of the measures studied so far, this minimum value is not independent of its parameter. For example consider,

$$\min_P H_a(P) = H_a(1, 0, 0, \dots, 0) = \frac{1}{a} \{ (1+a) \ln(1+a) - a \}.$$

which is definitely a positive quantity. And

$\lim_{a \rightarrow 0} \{ (1+a) \ln(1+a) - a \} = 0$. So if we consider $\min_a \{ \min_P H_a(P) \}$ then

it is zero. But for any other value of a , the minimum does not vanish. But this is undesirable, because when an outcome is certain to happen the uncertainty in that expt. should be zero.

But for this, the measures belonging to these four families satisfy most of the properties we have listed, of course additivity and subadditivity are not satisfied.

Measures of these four families are concave functions of p_1, p_2, \dots, p_n . Family (i) measures are concave with respect

to a, Family (ii) measures are convex w.r. to b, Family (iii) measures are also convex w.r. to c and Family (iv) measures are concave functions w.r. to k. with this we conclude discussion on these measures.

2.1.8 Behara and Chawla's Gamma Entropy :

It is given by $H_\gamma(P) = \frac{1}{(1-e^{1-\gamma})} \left\{ \sum_{i=1}^n p_i^{1/\gamma} - 1 \right\}^\gamma \neq 1, \gamma > 0$

$H_\gamma(P)$ is a non-negative continuous symmetric function of the probabilities. The minimum value occurs for degenerate distributions and is equal to zero. By applying Lagrange's multipliers method, we obtain that $U = (\frac{1}{n}, \dots, \frac{1}{n})$ is a constrained critical point for $H_\gamma(P)$ which can only correspond to local maximum as minimum is attained degenerate distributions. But whether this local maximum is also the global maximum depends on the concavity of the measure. However we are unable to conclude that $H_\gamma(P)$ is concave for any values of γ .

$H_\gamma(P)$ approaches Shannon's measure of information when $\gamma \rightarrow 1$. If γ is non-negative it satisfies the expansibility property.

For P and Q independent we have the following relation:

$$H_\gamma(PQ) = H_\gamma(P) + H_\gamma(Q) + (1-e^{\gamma-1}) H_\gamma(P)H_\gamma(Q)$$

from which we can easily see that $H_\gamma(P)$ is non-additive except when $\gamma = 1$.

$H_{\underline{\gamma}}(P)$ is not subadditive for all values of γ . This we achieve by providing the following example :

Example 2.1.8.1 : Let $P*Q = \begin{pmatrix} 0.02 & 0.08 \\ 0.27 & 0.63 \end{pmatrix}$ and $P = (0.1, 0.9)$ and $Q = (0.29, 0.71)$. Let $\gamma = 0.1$. Then we have

$$H_{\underline{0.1}}(P*Q) = 0.24001, H_{\underline{0.1}}(P) = 0.068511 \text{ and } H_{\underline{0.1}}(Q) = 0.193670$$

$$\text{and thereby } H_{\underline{0.1}}(P \ Q) = 0.24001 < 0.26781 = H_{\underline{0.1}}(P) + H_{\underline{0.1}}(Q).$$

For $\gamma = 0.1$, $H_{\underline{\gamma}}(P)$ is subadditive.

Again let $\gamma = 0.5$. Then we have

$$H_{\underline{0.5}}(P*Q) = 0.4773052, H_{\underline{0.5}}(P) = 0.14561 \text{ and } H_{\underline{0.5}}(Q) = 0.35925$$

$$\text{and so } H_{\underline{0.5}}(P*Q) = 0.4773052 < 0.50486 = H_{\underline{0.5}}(P) + H_{\underline{0.5}}(Q) \text{ again } H_{\underline{\gamma}} \text{ is subadditive.}$$

But consider $\gamma = 1.5$. Then we have

$$H_{\underline{1.5}}(P*Q) = 0.26746, H_{\underline{1.5}}(P) = 0.090511 \text{ and } H_{\underline{1.5}}(Q) = 0.146240$$

$$\text{and hence } H_{\underline{1.5}}(P*Q) = 0.26746 > 0.23675 = H_{\underline{1.5}}(P) + H_{\underline{1.5}}(Q)$$

and so $H_{\underline{\gamma}}$ is superadditive. For the distribution $P*Q$ we have chosen in this example, for $\gamma = 0.1$ and 0.5 , $H_{\underline{\gamma}}$ is subadditive but when we cross over to 1.5 , $H_{\underline{\gamma}}$ is not subadditive anymore. We shall discuss this property of $H_{\underline{\gamma}}$ in detail in Chap. 4.

Because of the very complicated expressions, we are not able to decide the monotonic behaviour of $H_{\underline{\gamma}}(P)$ w.r. to γ and also its concavity or convexity properties w.r. to γ . Again consider the following example.

Example 2.1.8.2 : Let $P * Q$, P and Q be as in Example 1.2. We get

$$H_{\underline{0.8}}(P * Q) = 0.108580, H_{\underline{0.8}}(P) = 0.037065 \text{ and } H_{\underline{0.8}}(Q) = 0.092797$$

$$H_{\underline{1.2}}(P * Q) = 1.1828908, H_{\underline{1.2}}(P) = 0.415840 \text{ and } H_{\underline{1.2}}(Q) = 0.720370$$

$$H_{\underline{2}}(P * Q) = 3.1944382, H_{\underline{2}}(P) = 0.94918, H_{\underline{2}}(Q) = 1.43568.$$

Now we can draw the following conclusions from the above data

$$(i) \quad H_{\underline{0.8}}(P * Q) < H_{\underline{0.1}}(P * Q) < H_{\underline{1.5}}(P * Q) < H_{\underline{0.5}}(P * Q)$$

$$< H_{\underline{1.2}}(P * Q) < H_{\underline{2}}(P * Q)$$

$$(ii) \quad H_{\underline{0.8}}(P) < H_{\underline{1.5}}(P) < H_{\underline{0.1}}(P) < H_{\underline{0.5}}(P) < H_{\underline{1.2}}(P) < H_{\underline{2}}(P) \text{ and}$$

$$(iii) \quad H_{\underline{0.8}}(Q) < H_{\underline{1.5}}(Q) < H_{\underline{0.1}}(Q) < H_{\underline{0.5}}(Q) < H_{\underline{1.2}}(Q) < H_{\underline{2}}(Q).$$

From the above relations we can easily conclude that $H_{\underline{\gamma}}(.)$ is not monotonic function of $\underline{\gamma}$ either for $0 < \underline{\gamma} < 1$ or $\underline{\gamma} > 1$.

We shall now conclude the discussion of Behara and Chawlas' Gamma entropy.

2.1.9 Sharma and Tanejas' Measures of Entropy

They are given as below :

$$(i) \quad H^{\alpha, \beta}(P) = (2^{1-\alpha} - 2^{1-\beta})^{-1} \left\{ \sum_{i=1}^n p_i^{\alpha} - \sum_{i=1}^n p_i^{\beta} \right\}, \alpha \neq \beta$$

$$(ii) \quad H^L(P) = -2^{\alpha-1} \left\{ \sum_{i=1}^n p_i^{\alpha} \ln p_i \right\}, \text{ and}$$

$$(iii) \quad H^S(P) = \frac{2^{\alpha-1}}{\sin \beta} \left\{ \sum_{i=1}^n p_i^{\alpha} \sin (\beta \log p_i) \right\}.$$

We shall consider these measures in the order given above

(i) Measure of order α and type β , $H^{\alpha,\beta}(P)$:

We can obviously see that $H^{\alpha,\beta}(P)$ is a non-negative, continuous and symmetric function of the probabilities. $H^{\alpha,\beta}(P)$ is also symmetric in α, β . Kapur [48] has thoroughly investigated the range of validity of this measure. He has established that for this measure the maximum may not always occur for the uniform distribution. His result states that if

$$\frac{1}{(\beta-\alpha)} \left\{ \frac{\alpha(\alpha-1)}{n^\alpha} - \frac{\beta(\beta-1)}{n^\beta} \right\} > 0$$

then $H^{\alpha,\beta}(\cdot)$ has a local minimum at $U = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$. However he has also proved that [31] if $\alpha < 1$ and $\beta > 1$ or $\alpha > 1$ and $\beta < 1$, $H^{\alpha,\beta}(P)$ is a concave measure and the maximum occurs for U . The minimum value is zero and it occurs for degenerate distributions.

The maximum of $H^{\alpha,\beta}(P)$ for $\alpha < 1$ and $\beta > 1$ or vice versa is given by

$$\max_P H^{\alpha,\beta}(P) = H^{\alpha,\beta}(U) = \frac{n^{1-\alpha} - n^{1-\beta}}{2^{1-\alpha} - 2^{1-\beta}}.$$

and we can easily see that $H^{\alpha,\beta}(U)$ is an increasing function of n .

Now we shall consider the additivity and subadditivity properties of $H^{\alpha,\beta}(P)$. It is not additive. For, if we take $\beta=1$, $H^{\alpha,1}(P) = H^\alpha(P)$, the Havrda-Charvat's measure of entropy which was shown to be non-additive.

$H^{\alpha,\beta}(P)$ is also not subadditive for all values of α and β . We can give here just one example which verifies our claim. But a detailed discussion on this property is included in Chap. 4.

Example 2.1.9.1 : Let $\alpha = 0.5$ and $\beta = 1.5$ and $P * Q$, P and Q be as in Example 1.2. Then we have

$$H^{0.5,1.5}(P * Q) = 1.5157705, H^{0.5,1.5}(P) = 0.53665 \text{ and}$$

$$H^{0.5,1.5}(Q) = 0.88629.$$

Therefore $H^{0.5,1.5}(P * Q) = 1.5157705 > 1.42284 = H^{0.5,1.5}(P) + H^{0.5,1.5}(Q)$. Hence for this set of values $H^{\alpha,\beta}(P)$ is not subadditive.

Now we shall discuss the properties of this measure as function of its parameters.

Now we shall prove the following result regarding concavity and convexity of $H^{\alpha,\beta}(P)$ w.r. to α and β .

Proposition 2.1.9.1 : $H^{\alpha,\beta}(P)$ is

- (i) Convex w.r.to α if $\alpha < 1$ and $\beta > 1$
- (ii) Concave w.r.to α if $\alpha > 2$ and $\beta < 1$
- (iii) Concave w.r.to β if $\alpha < 1$ and $\beta > 2$ and
- (iv) Convex w.r.to β if $\alpha > 1$ and $\beta < 1$.

Proof : Let $f(\alpha) = (2^{1-\alpha} - 2^{1-\beta})^{-1} \left(\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta \right)$. Then we have

$$f'(\alpha) = \frac{\sum_{i=1}^n p_i^\alpha \log p_i}{(2^{1-\alpha} - 2^{1-\beta})} + \frac{2^{1-\alpha} (\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta)}{(2^{1-\alpha} - 2^{1-\beta})^2} \text{ and}$$

$$f''(\alpha) = \frac{\sum_{i=1}^n p_i^\alpha (\log p_i)^2}{(2^{1-\alpha} - 2^{1-\beta})} + \frac{2^{2-\alpha} \sum_{i=1}^n p_i^\alpha \log p_i}{(2^{1-\alpha} - 2^{1-\beta})^2}$$

$$- \frac{2^{1-\alpha} (\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta) (1 - 2^{2-\alpha})}{(2^{1-\alpha} - 2^{1-\beta})^3}$$

- (i) $\alpha < 1$ and $\beta > 1$. We recall the arguments of 1.3.2 and conclude that the sum of the first two terms in $f''(\alpha)$ is positive. Then we know that $\{(\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta) / (2^{1-\alpha} - 2^{1-\beta})^3\}$ is always positive. $(1 - 2^{2-\alpha})$ is negative because $\alpha < 1$. Therefore we get that $f''(\alpha) > 0$. That is $f(\alpha)$ is convex.
- (ii) $\alpha > 2$, $\beta < 1$. Then we have obviously the first two terms of $f''(\alpha)$ negative. Again $\{(\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta) / (2^{1-\alpha} - 2^{1-\beta})^3\}$ is positive. But now $(1 - 2^{2-\alpha})$ is also positive making the third term of $f''(\alpha)$ also negative. Therefore we now have $f''(\alpha) < 0$. That is $f(\alpha)$ is concave.

Now let $g(\beta) = (2^{1-\alpha} - 2^{1-\beta})^{-1} (\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta)$. Then we have

$$g'(\beta) = \frac{-\sum_{i=1}^n p_i^\beta \log p_i}{(2^{1-\alpha} - 2^{1-\beta})} - \frac{2^{1-\beta} (\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta)}{(2^{1-\alpha} - 2^{1-\beta})^2} \text{ and}$$

$$g''(\beta) = \frac{-\sum_{i=1}^n p_i^\beta (\log p_i)^2}{(2^{1-\alpha} - 2^{1-\beta})} - \frac{2^{1-\beta} (\sum_{i=1}^n p_i^\alpha - \sum_{i=1}^n p_i^\beta) (1 - 2^{2-\beta})}{(2^{1-\alpha} - 2^{1-\beta})^3}$$

- (iii) Let $\alpha < 1$ and $\beta > 2$. Then we can easily see that $(2^{1-\alpha} - 2^{1-\beta}) > 0$ and so the first term in $g''(\beta)$ is negative and by the same arguments as in (i) and (ii) above, the second term is also negative. Therefore $g''(\beta) < 0$ and $g(\beta)$ is concave.
- (iv) Let $\alpha > 1$ and $\beta < 1$ with similar arguments as above we obtain that $g''(\beta) > 0$ and so $g(\beta)$ is convex. It completes the proof.

But we are unable to decide the nature of $f(\alpha)$ and $g(\beta)$ in the cases (a) $\beta < 1$, $1 < \alpha < 2$ and (b) $\alpha < 1$, $1 < \beta < 2$.

Although we could decide about the sign of the second derivative of $f(\alpha)$ and $g(\beta)$ we are unable to do so for $f'(\alpha)$ and $g'(\beta)$. We know that $\beta = 1$, $H^{\alpha, \beta}(\cdot)$ reduces to Havrda and Charvat's measure, which is a monotonically decreasing function of α with that we conclude our discussion of this measure.

(ii) Logarithmic Measure of Entropy :

$H_L(P)$ is actually a limiting case of $H^{\alpha, \beta}(P)$ as $\beta \rightarrow \alpha$. We can easily verify that. It is also easy to see that $H_L(P)$ is a continuous symmetric function of the probabilities. $H_L(P)$ is always non-negative. And $H_L(P)$ vanishes for any degenerate distributions and hence the minimum value of it is zero.

Kapur [41] had established for $U = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ has a local minimum if $\alpha > 1$ and $n > \exp \{(2\alpha-1)/\alpha(\alpha-1)\}$. But once we take the care that these inequalities are not satisfied, we have

$H_L(P)$ being maximum at $P = U$. We have

$$\max_P H_L(P) = H_L(U) = 2^{\alpha-1} n^{1-\alpha} \log n.$$

Let us denote by $f(n) = H_L(U)$. We then have

$$f'(n) = 2^{\alpha-1} n^{-\alpha} \{(1-\alpha) \log n + 1\};$$

$$f'(n) > 0 \iff (1-\alpha) \log n > -1 \text{ or } n > 2^{\frac{1}{\alpha-1}} \text{ if } 0 < \alpha < 1$$

$$n < 2^{\frac{1}{\alpha-1}} \text{ if } \alpha > 1.$$

Therefore the maximum of $H_L(P)$ is an increasing function of n iff it satisfies the above conditions. For example if $\alpha = 0.5$ then n should be at least 5 and if $\alpha = 1.5$ then n can not be greater than 4, in order that $f(n)$ be an increasing function of n .

Now we shall consider additivity and subadditivity of $H_L(P)$.

Additivity : For P and Q independent and PQ being their product distribution we get the relation

$$H_L(PQ) = \left(\sum_{j=1}^m q_j^\alpha \right) H_L(P) + \left(\sum_{i=1}^n p_i^\alpha \right) H_L(Q)$$

Therefore we can see that $H_L(P)$ is not additive in general unless $\alpha = 1$ in which case $H_L(P)$ reduces to Shannon's measure.

Subadditivity : We prove the following result on the subadditivity of $H_L(P)$.

Proposition 2.1.9.2 : $H_L(.)$ is subadditive if $\alpha > 1$ and not subadditive for $\alpha < 1$.

Proof : Let $\{\pi_{ij}\}_{i=1, \dots, n}^{j=1, \dots, m}$ be a joint probability distribution and let $P = \{p_i\}_i^n$ and $Q = \{q_j\}_j^m$ be the corresponding marginal probability distributions. We now have

$$\pi_{ij} = p_i q_{ji}$$

where q_{ji} is the conditional probability that j^{th} event of the second distribution occurs when it is known that i^{th} event of the first distribution occurred. Now let $\alpha > 1$, we have

$$\begin{aligned} H_L(P * Q) &= -2^{\alpha-1} \sum_{i=1}^n \sum_{j=1}^m (p_i q_{ji})^\alpha \log (p_i q_{ji}) \\ &= -2^{\alpha-1} \sum_{i=1}^n p_i^\alpha \log p_i \left(\sum_{j=1}^m q_{ji}^\alpha \right) - 2^{\alpha-1} \sum_{i=1}^n p_i^\alpha \sum_{j=1}^m q_{ji}^\alpha \log q_{ji} \\ &\leq -2^{\alpha-1} \sum_{i=1}^n p_i^\alpha \log p_i - 2^{\alpha-1} \sum_{i=1}^n p_i^\alpha \sum_{j=1}^m p_i^\alpha \sum_{j=1}^m q_{ji}^\alpha \log q_{ji} \\ &\quad \left[\text{because } \sum_{j=1}^m q_{ji}^\alpha < \sum_{j=1}^m q_{ji} = 1 \right] \\ &= H_L(P) + \sum_{i=1}^n p_i^\alpha H_L(Q/P_i) \\ &\leq H_L(P) + H_L(Q) \quad \text{just as in the case of Shannon's measure.} \end{aligned}$$

This proof does not hold in the case of $\alpha < 1$ because if

$$\alpha < 1, \quad \sum_{i=1}^n p_i^\alpha > 1.$$

Example 2.1.9.2 : Let P and Q and $(P * Q)$ be the same as the distributions taken in Example 2.1.4.1

- (i) $\alpha = 0.5$: We have $H_L(P * Q) = 0.710829$, $H_L(P) = 0.2543018$ and $H_L(Q) = 0.2933357$. Therefore we have
- $$H_L(P * Q) = 0.7108229 > 0.5476375 = H_L(P) + H_L(Q).$$

(ii) $\alpha = 2$: We now have $H_L(P * Q) = 0.2575902$, $H_L(P) = 0.0941271$,
 $H_L(Q) = 0.2403859$

Therefore $H_L(P * Q) = 0.2575902 < 0.334513 = H_L(P) + H_L(Q)$.

The results in this example are obviously consistent with those in Proposition 2.1.9.2.

Now we shall finally consider the properties of $H_L(P)$ as a function of its parameter.

Proposition 2.1.9.3 : Let $f(\alpha) = -2^{\alpha-1} \sum_{i=1}^n p_i^\alpha \log p_i$. Then $f(\alpha)$ is monotonically decreasing and convex functions of α .

Proof : We have from the definition

$$f'(\alpha) = -2^{\alpha-1} \sum_{i=1}^n p_i^\alpha \{\log p_i + (\log p_i)^2\} \text{ and}$$

$$f''(\alpha) = -2^{\alpha-1} \sum_{i=1}^n p_i^\alpha \log p_i \{1 + \log p_i + 2(\log p_i)^2\}.$$

Obviously, $(\log p_i)^2$ is greater than $|\log p_i|$ and therefore we get that $f'(\alpha) < 0$ and $f''(\alpha) > 0$. i.e., $f(\alpha)$ is a monotonically decreasing convex function of α . That completes the proof.

With that we conclude the discussion on Sharma and Taneja's logarithmic measure of entropy.

(iii) Sine Measure of Entropy :

This measure does not satisfy most of the required properties of measures of entropy listed in 2.1.2.

The foremost drawback of $H_S(P)$ is that it can be negative also. For consider the following

Example 2.1.9.3 : Let $\alpha = 2$, $\beta = -1.348$, $n = 2$, $P = (0.2, 0.8)$.

Then we have

$$H_S(P) = -0.5708569 < 0.$$

Maximum and Minimum :

Even though $H_S(0, 1, 0, \dots, 0) = 0$ we can not claim that zero is the minimum value of $H_S(P)$ because $H_S(P)$ can also assume negative values. Kapur [41] has discussed about $H_S(P)$ being negative. But we are not able to obtain the absolute minimum value $H_S(P)$ attains.

We can not also prove that $H_S(P)$ is maximum when $P = U$ or if the value $H_S(U)$ is an increasing function of n . We can not prove or disprove any of the other properties like concavity with respect to P monotonicity and concavity (convexity) with respect to either of the parameters, all due to the extremely complicated expressions.

However we can conclude that $H_S(P)$ is neither additive nor subadditive.

Example 2.1.9.4 : Let $\alpha = 2$, $\beta = 1.5$, $P = (0.5, 0.5)$, $Q = (0.4, 0.6)$, $P * Q = PQ$, the product distribution. Then we have $H_S(PQ) = -0.131$, $H_S(P) = -0.999$ and $H_S(Q) = -0.938$. Therefore we get that $H_S(P*Q) = -0.131 > -1.937 = H_S(P) + H_S(Q)$, violating the subadditivity law.

2.1.10 Rathie's Measure of Entropy with $(n+1)$ parameters :

$$\text{It is given by } H_{\alpha}^{\beta_1, \beta_2, \dots, \beta_n}(P) = \frac{1}{2^{1-\alpha-1}} \left[\frac{\sum_{i=1}^n p_i^{\alpha+\beta_i-1}}{\sum_{i=1}^n p_i^{\beta_i}} - 1 \right], \alpha \neq 1.$$

Note that if we take $\beta_1 = \beta_2 = \dots = \beta_n = 1$, we get Havrda-Charvat measure of entropy. This measure is not a symmetric function of the probabilities. However it is a non-negative continuous function of the parameters.

We shall now consider the maximum and minimum values of this measure. Obviously the minimum value is zero which is attained for any degenerate distribution.

We shall by way of an example show that Rathie's measure does not always attain its maximum for uniform distribution.

Example 2.1.10.1 : Let $n = 2$, $\beta_1 = \frac{1}{2}$, $\beta_2 = 1$, $\alpha = 2$ and $P_1 = (0.5, 0.5)$ and $P_2 = (0.4, 0.6)$. Then we have $H_2^{0.5, 1}(P_1) = 1.0$ and $H_2^{0.5, 1}(P_2) = 1.0052668$. Therefore we have $H_2^{0.5, 1}(P_1) < H_2^{0.5, 1}(P_2)$. So this example justifies our claim that $H_{\alpha}^{\beta_1, \dots, \beta_n}(P)$ does not always attain its maximum value for uniform distributions. And this is obvious because of the non-symmetry of $H_{\alpha}^{\beta_1, \dots, \beta_n}(P)$ w.r.to P .

There is a problem in proceeding for either proving or disproving additivity and subadditivity for Rathie's measure. Because in $P * Q$ there are mn components, we need mn parameters

$\{\beta_i\}$ besides α . But we only have n parameters from the definition of the measure. Unless this question is resolved the additivity and subadditivity properties are not defined for this measure of entropy. In fact it is a serious draw back as we often need to measure the entropy of a joint probability distribution as well as the marginal entropies.

Also the expansibility property is not defined for this measure because of the same problem as above stated.

Because $H_{\alpha}^{\beta_1, \beta_2, \dots, \beta_n(P)}$ does not have the sum property which is possessed by each of the measures considered so far, we are unable to put forth any arguments for either proving or disproving the concavity of $H_{\alpha}^{\beta_1, \beta_2, \dots, \beta_n(P)}$ with respect to P .

We shall now consider the monotonic behaviour of it w.r.to the parameters.

Proposition 2.1.10.1 : $H_{\alpha}^{\beta_1, \dots, \beta_n(P)}$ is a monotonically increasing function of α for $\alpha > 1$. For $0 < \alpha < 1$, it is not a monotonic function of α .

Proof : Let $f(\alpha) = H_{\alpha}^{\beta_1, \dots, \beta_m(P)}$. Then we have

$$f'(\alpha) = \frac{(2^{1-\alpha}-1) \left(\sum_{i=1}^n p_i^{\alpha+\beta_i-1} \log_2 p_i \right) + \left(\sum_{i=1}^n p_i^{\alpha+\beta_i-1} \right) 2^{1-\alpha}}{(2^{1-\alpha}-1)^2 \left(\sum_{i=1}^n p_i^{\beta_i} \right)} + \frac{2^{1-\alpha}}{(2^{1-\alpha}-1)^2}.$$

Now let $\alpha > 1$ then we have $(2^{1-\alpha}-1) < 0$, $\log p_i \leq 0 \forall i=1,2,\dots,n$ and therefore we get that $f'(\alpha) > 0$.

But for $0 < \alpha < 1$ we can not arrive at a definite conclusion as we can not estimate which of $\left| \sum_{i=1}^n p_i^{\alpha+\beta_i-1} \log p_i \right|$ and $\sum_{i=1}^n p_i^{\alpha+\beta_i-1}$ is greater. Therefore no conclusion about $f'(\alpha)$ for $0 < \alpha < 1$ can be drawn.

Now if we consider $g(\beta_i) = H_{\alpha}^{\beta_1, \dots, \beta_n}(p)$ we then have

$$g'(\beta_k) = \frac{1}{(2^{1-\alpha}-1)} \frac{p_k^{\beta_k} \log p_k (p_k^{\alpha-1} \sum_{i=1}^n p_i^{\beta_i} - \sum_{i=1}^n p_i^{\alpha+\beta_i-1})}{(\sum_{i=1}^n p_i^{\beta_i})^2}, \quad 1 \leq k \leq n$$

But here also we face the same dilemma about the relative magnitudes of $(p_k^{\alpha-1} \sum_{i=1}^n p_i^{\beta_i})$ and $\sum_{i=1}^n p_i^{\alpha+\beta_i-1}$.

With that we conclude the discussion of Rathie's measure of entropy with $(n+1)$ parameters.

2.1.11 Van der Lubbe et al., Measures of Entropy : They are given by

$$\begin{aligned} \text{(i)} \quad H_n^1(p; \rho, \sigma, \delta) &= -\delta \log \left[\sum_{i=1}^n p_i^{\rho} \right]^{\sigma} \\ \text{(ii)} \quad H_n^2(p; \rho, \sigma, \delta) &= \delta \left[1 - \left(\sum_{i=1}^n p_i^{\rho} \right)^{\sigma} \right] \quad \text{and} \\ \text{(iii)} \quad H_n^3(p; \rho, \sigma, \delta) &= \delta \left[\left(\sum_{i=1}^n p_i^{\rho} \right)^{-\sigma} - 1 \right] \end{aligned}$$

where $(\rho, \sigma) \in D = \{(\rho, \sigma) | 0 < \rho < 1, \sigma < 0 \vee \rho > 1, \sigma > 0\}$ and δ , a positive normalizing constant.

Unlike the classical approach where measures of entropy are derived by characterizing the uncertainty of a random variable, Van der Lubbe et al. [61] characterized the

certainly of a random variable and then considered monotonic decreasing functions of this certainty measure for measures of entropy.

The following are the axioms assumed by Lubbe et al. for obtaining the measures stated above : [Thm. 4 [61]] .

(1) For stochastically independent experiments X and Y it holds that

$$H(PQ) = H(P) + H(Q) + c H(P) \cdot H(Q), \quad c \in \mathbb{R}.$$

(2) The information measure $H(P)$ is non-negative and a continuous and strictly monotonic function of the certainty measure $f(P)$.

Then they proved that the only non-trivial solutions $H(P)$ are as follows :

(A) For $c = 0$, $H_n^1(P; \rho, \sigma, \delta)$

(B) For $c < 0$, $H_n^2(P; \rho, \sigma, \delta)$ and

(C) For $c > 0$, $H_n^3(P; \rho, \sigma, \delta)$.

So it is evident that only the measures belonging to family (i) are additive. The measures belonging to other two families are non-additive.

They also obtained the following results :

(1) [Corollary 4 of [61]] $\max_i H_n^1(P; \rho, \sigma, \delta) = -(1-\rho)\sigma \delta \log_2 n$

$$\max_i H_n^2(P; \rho, \sigma, \delta) = \delta [1 - n^{-(1-\rho)\sigma}]$$

$$\max_i H_n^3(P; \rho, \sigma, \delta) = \delta [n^{-(1-\rho)\sigma} - 1]$$

and $\min_j H_n^j(P; \rho, \sigma, \delta) = H_n^j(1, 0, 0, \dots, 0; \rho, \sigma, \delta) = 0$. $j = 1, 2, 3$.

(2) [Corollary 5 of [61]] (i) For fixed σ and δ it holds that

$H_n^i(P; \rho, \sigma, \delta)$, $i = 1, 2, 3$, are

increasing w.r.to ρ for $\rho > 1$, $\sigma > 0$, $\delta > 0$ and

decreasing w.r. to ρ for $0 < \rho < 1$, $\sigma < 0$, $\delta > 0$.

(ii) For fixed ρ and δ it holds that $H_n^i(P; \rho, \sigma, \delta)$, $i = 1, 2, 3$ are

increasing w.r. to σ for $\rho > 1$, $\sigma > 0$ and $\delta > 0$ and

decreasing w.r.to σ for $0 < \rho < 1$, $\sigma < 0$ and $\delta > 0$.

Now we proceed to prove the following results :

Proposition 2.1.11.1 :

(i) $H_n^1(P; \rho, \sigma, \delta)$ is a concave function of ρ for $0 < \rho < 1$ and $\sigma < 0$,

(ii) There exists a σ_1 s.t. for $\rho > 1$ and $\sigma \leq \sigma_1$, $H_n^1(P; \rho, \sigma, \delta)$ is concave with respect to ρ and for $\sigma \geq \sigma_1$ it is convex w.r.to ρ ,

(iii) $H_n^1(P; \rho, \sigma, \delta)$ is both concave and convex w.r.to σ , for $(\rho, \sigma) \in D$.

Proof : Let $f_1(\rho) = H_n^1(P; \rho, \sigma, \delta)$. Then we have

$$f_1'(\rho) = -\delta \left[\sum_{i=1}^n p_i \right]^{-\sigma} \sum_{i=1}^n p_i \log p_i \text{ and}$$

$$f_1''(\rho) = -\delta \left[\sum_{i=1}^n p_i \right]^{-(\sigma+1)} \left[-\sigma \left(\sum_{i=1}^n \frac{p_i^\rho}{\sum p_i^\rho} \log p_i \right)^2 + \sum_{i=1}^n \frac{p_i^\rho}{\sum p_i^\rho} (\log p_i)^2 \right]$$

(i) Let $0 < \rho < 1$ and $\sigma < 0$. Then we can see immediately that

$f_1''(\rho) < 0$ and therefore $H_n^1(P; \rho, \sigma, \delta)$ is concave w.r. to

in this case.

(ii) Now let $\rho > 1$ and $\sigma > 0$. Then we have from the convexity of $x \rightarrow x^2$

$$\left(\sum_{i=1}^n \frac{p_i^\rho}{\sum_{i=1}^n p_i^\rho} \log p_i \right)^2 \leq \sum_{i=1}^n \frac{p_i^\rho}{\sum_{i=1}^n p_i^\rho} (\log p_i)^2$$

$$+ \sum_{i=1}^n \frac{p_i^\rho}{\sum_{i=1}^n p_i^\rho} (\log p_i)^2$$

Now let $\sigma_1 = \frac{\sum_{i=1}^n \frac{p_i^\rho}{\sum_{i=1}^n p_i^\rho} (\log p_i)^2}{\sum_{i=1}^n \frac{p_i^\rho}{\sum_{i=1}^n p_i^\rho} \log p_i)^2}$. Then we can see easily

that $|\sigma - \left(\sum_{i=1}^n \frac{p_i^\rho}{\sum_{i=1}^n p_i^\rho} \log p_i \right)^2| \lesssim \left| \sum_{i=1}^n \frac{p_i^\rho}{\sum_{i=1}^n p_i^\rho} (\log p_i)^2 \right|$ depending on

$\sigma \lesssim \sigma_1$. Therefore $f_1''(\rho) \lesssim 0$ according as $\sigma \lesssim \sigma_1$. If $\sigma = \sigma_1$ then we have that $f_1''(\rho) = 0$. That is $H_n^1(P; \rho, \sigma, \delta)$ is both concave and convex.

(iii) Now let $g_1(\sigma) = H_n^1(P; \rho, \sigma, \delta)$. We then have

$$g_1'(\sigma) = - \log_2 \sum_{i=1}^n p_i \quad \text{and} \quad g_1''(\sigma) \equiv 0 \text{ which completes}$$

the proof of Proposition 2.1.11.1.

Proposition 2.1.11.2 :

- (i) $H_n^2(P; \rho, \sigma, \delta)$ is a concave function for $\sigma \geq \sigma_2$ and a convex function $\sigma \leq \sigma_2$ w.r. to ρ where $(\rho, \sigma) \in D$ and $\sigma_2 = (1 - \sigma_1)$
- (ii) $H_n^2(P; \rho, \sigma, \delta)$ is a concave function w.r. to σ for $(\rho, \sigma) \in D$.

Proof : (i) Let $f_2(\rho) = H_n^2(P; \rho, \sigma, \delta)$. Then we have

$$f_2'(\rho) = -\delta \sigma \left(\sum_{i=1}^n p_i^\rho \right)^{\sigma-1} \sum_{i=1}^n p_i^\rho \log p_i \text{ and}$$

$$f_2''(\rho) = -\delta \sigma \left[\sum_{i=1}^n p_i^\rho \right]^\sigma \left[\left(\sum_{i=1}^n \frac{p_i^\rho}{\sum p_i^\rho} \log p_i \right)^2 (\sigma-1) \right. \\ \left. + \sum_{i=1}^n \frac{p_i^\rho}{\sum p_i^\rho} (\log p_i)^2 \right].$$

$$\text{Now let } \sigma_2 = 1 - \sigma_1 = 1 - \frac{\sum p_i^\rho (\log p_i)^2}{\left(\sum \frac{p_i^\rho}{\sum p_i^\rho} \log p_i \right)^2}.$$

Now we can clearly see that for $\sigma = \sigma_2$ $f_2''(\rho) = 0$ and for $\sigma \leq \sigma_2$ $f_2''(\rho) \geq 0$ and for $\sigma \geq \sigma_2$, $f_2''(\rho) \leq 0$, again making use of the convexity of the function $x \rightarrow x^2$.

Note here that $\sigma_2 < 0$ and $\sigma_1 > 0$. Therefore for all $\rho > 1$ and $(\rho, \sigma) \in D$ $H_n^2(P; \rho, \sigma, \delta)$ is a concave function of ρ and there corresponds point ρ_2 to σ_2 such that for $0 < \rho \leq \rho_2$, $H_n^2(P; \rho, \sigma, \delta)$ is convex and for $\rho \geq \rho_2$ it concave w.r. to ρ .

(ii) Let $g_2(\sigma) = H_n^2(P; \rho, \sigma, \delta)$. Then we have

$$g_2'(\sigma) = -\delta \left[\sum_{i=1}^n p_i^\rho \right]^\sigma \log \left[\sum p_i^\rho \right] \text{ and}$$

$$g_2''(\sigma) = -\delta \left[\sum_{i=1}^n p_i^\rho \right]^\sigma \left[\log \sum_{i=1}^n p_i^\rho \right]^2$$

We can clearly see that $g_2''(\sigma) \leq 0$ for all $(\rho, \sigma) \in D$. That completes the proof of Proposition 2.1.11.2.

Proposition 2.1.11.3 :

- (i) $H_n^3(P; \rho, \sigma, \delta)$ is a concave function w.r. to ρ for $\sigma \leq \sigma_3$ and a convex function w.r. to ρ for $\sigma \geq \sigma_3$ where $\sigma_3 = \sigma_1 - 1$.
- (ii) $H_n^3(P; \rho, \sigma, \delta)$ is a convex function w.r. to σ for all $(\rho, \sigma) \in D$.

Proof : (i) It is exactly similar to that (i), Proposition 2.1.11.2 thus we omit it here. We note, however, that $\sigma_3 > 1$. Therefore we have the relation :

$$\sigma_2 < 0 < \sigma_3 < \sigma_1.$$

- (ii) Let $g(\sigma) = H_n^3(P; \rho, \sigma, \delta)$. Then we have

$$g'(\sigma) = -\delta \left[\sum_{i=1}^n p_i^\rho \right]^{-\sigma} \log \sum_{i=1}^n p_i^\rho \quad \text{and}$$

$$g''(\sigma) = \delta \left[\sum_{i=1}^n p_i^\rho \right]^{-\sigma} (\log \sum_{i=1}^n p_i^\rho)^2$$

Therefore $g''(\sigma) \geq 0$ for all $(\rho, \sigma) \in D$. That completes the proof of Proposition 2.1.11.2.

We have now completed the study of the properties of Lubbe's measures of entropy w.r. to their parameters. Now we shall deal with the concavity of these measures w.r. to $P \in A_n$. we have the following

Proposition 2.1.11.4 : $H_n^i(P; \rho, \sigma, \delta)$ is pseudo-concave, concave and pseudo-concave for $i = 1, 2, 3$ respectively for $(\delta, \sigma) \in D_c = \{\rho, \sigma \mid 0 < \rho < 1, \sigma < 0 \vee \rho > 1, \sigma \geq 1\}$.

Proof : Theorem 2 of [61] states that for $(\rho, \sigma) \in D_c$, the certainty measure $\left[\sum_{i=1}^n p_i^\rho \right]^\sigma$ is convex with respect to $P \in A_n$.

Now making use of the following results, [15] we get our results :

- (i) \log (convex function) is pseudo-convex
- (ii) Negative of a convex (pseudo-convex) function is a concave (pseudo-concave) function and vice versa.
- (iii) Concave function/convex function is a pseudo-concave function.

That completes the proof of Proposition 2.1.11.4. Now we shall consider the subadditivity property of these measures. None of the three measures is subadditive. Consider the following examples :

Example 2.1.11.1 : Let $\delta = 1$, $\rho = 2$, $\sigma = 2$, $P * Q$, P and Q are same as in Example 2.1.4.1. We then have

$$H_4^1(P * Q; 2, 2, 1) = 1.4821554, H_2^1(P; 2, 2, 1) = 0.3969018 \text{ and}$$

$$H_2^1(Q; 2, 2, 1) = 1.0613765.$$

Therefore $H_4^1(P * Q; 2, 2, 1) > H_2^1(P; 2, 2, 1) + H_2^1(Q; 2, 2, 1)$.

Example 2.1.11.2 : Let $\delta = 1$, $\rho = 2$, $\sigma = 0.05$, P and Q as in Example 2.1.4.1. Then we have

$$H_4^2(P * Q; 2, 0.05, 1) = 0.0363757, H_2^2(P; 2, 0.05, 1) = 0.0261854 \text{ and}$$

$$H_2^2(Q; 2, 0.05, 1) = 0.0098734$$

Therefore $H_4^2(P * Q; 2, 0.05, 1) > H_2^2(P; 2, 0.05, 1) + H_2^2(Q; 2, 0.05, 1)$.

Example 2.1.11.3 : Let $\delta = 1$, $\rho = 2$, $\sigma = 0.05$, $P = (0.5, 0.5)$, $Q = (0.4, 0.6)$ and $P * Q = PQ$ the product distribution. Then we have

$$H_4^3(P * Q; 2, 0.05, 1) = 0.0696737, H_2^3(P; 2, 0.06, 1) = 0.0352649 \text{ and}$$

$$H_2^3(Q; 2, 0.05, 1) = 0.0332367$$

$$\text{Therefore } H_4^3(P * Q; 2, 0.05, 1) > H_2^3(P; 2, 0.05, 1) + H_2^3(Q; 2, 0.05, 1).$$

Therefore we had seen that the measures belonging to each of the three families proposed by Lubbe satisfy all the properties of entropy measures we listed in 2.1.2, except the additivity and subadditivity properties. With that we complete our study of Lubbe et al.'s measures of entropy.

2.1.12 Sharma and Mittals' Measure of Entropy :

It is given by $H_{a,b}(P) = \frac{1}{(1-a)^b} \{ (\sum_{i=1}^n p_i^a)^b - 1 \}$ $a > 0$,

$a \neq 1$, $b \neq 0$. $H_{a,b}(P)$ is a generalization of $H^a(P)$ which we will obtain by taking $b = 1$. Kapur has discussed the validity of $H_{a,b}(P)$ in [30]. He has established that

- (i) $H_{a,b}(P)$ is always non-negative
- (ii) Min of $H_{a,b}(P)$ is zero and is attained by degenerate distributions
- (iii) Max of $H_{a,b}(P)$ always occurs for the uniform distribution.
- (iv) $H_{a,b}(P)$ is non-additive.
- (v) $H_{a,b}(P)$ is concave w.r. to P for $a > 1 \forall b > 1$ or $0 < a < 1 \forall 0 < b < 1$ or for $a < 1, b < 0$.
- (vi) $H_{a,b}(P)$ is a monotonic decreasing function of $a \forall b$.

We can not conclude the monotonic behaviour of $H_{a,b}(P)$ w.r. to b and concavity or convexity of $H_{a,b}(P)$ is also equally complicated. This measure is of interest because it generalizes

Havrda-Charvat's, Renyi's and Shannon's measures of entropy without loosing many of their properties. The only properties lost are additivity (valid for Shannon and Renyi measures only) and subadditivity which is true always for only Shannon's measure and concavity with respect to P for some values of a and b, for example $a > 1$ and $0 < b < 1$ or $0 < a < 1$, $b > 1$.

With that we close our discussion of Sharma and Mittals' measure of entropy.

2.1.13 Kapur's Generalized Measures of Entropy : The measures we consider in this subsection are given by

$$(i) \quad H_{\alpha, \beta}^b(P) = \frac{(\sum_{i=1}^n p_i^\alpha)^b - (\sum_{i=1}^n p_i^\beta)^b}{(\beta - \alpha)b} \quad \alpha \neq \beta, b \neq 0, \alpha, \beta > 0$$

$$(ii) \quad H_k^1(P) = \frac{1}{(\alpha - \beta)} \left\{ \left[\left(\frac{\lambda}{n} + (1 - \lambda)p_1 \right)^{1 - \alpha} - \sum_{i=1}^n \left(\frac{\lambda}{n} + (1 - \lambda)p_i \right)^{1 - \beta} \right] \right. \\ \left. - \sum_{i=1}^n p_i^\alpha \left(\frac{\lambda}{n} + (1 - \lambda)p_i \right)^{1 - \alpha} + \sum_{i=1}^n p_i^\beta \left(\frac{\lambda}{n} + (1 - \lambda)p_i \right)^{1 - \beta} \right\}$$

$$(iii) \quad H_k^2(P) = \frac{1}{(\alpha - 1)a} \left\{ \left(\frac{\lambda}{n} + (1 - \lambda) \right)^{(1 - \alpha)a} - \sum_{i=1}^n p_i^\alpha \left(\frac{\lambda}{n} + (1 - \lambda)p_i \right)^{(1 - \alpha)a} \right\}$$

$$(iv) \quad H_k^3(P) = \frac{1}{(\alpha - 1)} \ln \frac{\left\{ \frac{\lambda}{n} + (1 - \lambda) \right\}^{1 - \alpha}}{\sum_{i=1}^n p_i^\alpha \left(\frac{\lambda}{n} + (1 - \lambda)p_i \right)^{1 - \alpha}}.$$

Now we shall consider the properties of these measures.

- (i) The important feature of this measure is that it generalizes many known measures of entropy. For $b = 1$ we get Sharma and Taneja's measure of order α and type β , for $b \rightarrow 0$

we get Kapur, Aczel and Daroczy's measure of order α and type β , and other measures like Renyi's, Havrda-Charvats', Sharma and Mittals' and Shannon's are all special cases of this measure.

Except for the additivity, concavity w.r. to both P and the parameters and the subadditivity, $H_{\alpha, \beta}^a(P)$ satisfies all other properties which we have listed in 1.2.

The additivity is satisfied for the special cases when Shannon's or Renyi's or Kapur-Aczel Daroczys' measures result.

- (ii) Based on the definition $H(P) = \max_P D(P:U) - D(P:U)$, Kapur has given a systematic method of obtaining entropy measures from valid Directed Divergence measures. This three parametric measure $H_K^1(P)$ is thus obtained by constructing a three parametric directed divergence measure :

$$D(P:Q) = \sum_{i=1}^n q_i \left(f\left(\frac{p_i}{q_i}\right) - f(1) \right)$$

where $f(x)$ is a twice differentiable convex function.

By considering special cases for $f(x)$ in $D(P;Q)$ and the above definition for $H(P)$, Kapur [37] obtained

$H_K^1(P)$, $H_K^2(P)$ and $H_K^3(P)$.

All these measures, by their definition satisfy all the basic properties of measures of information, like non-negativity continuity and symmetry with respect to the

probabilities, minimum for degenerate distributions and max. for uniform distribution and the maximum value is an increasing function of n . They also satisfy the expansibility property.

With that we conclude our discussion on measures or entropy and close this section with a table with properties listed in 2.1.2 against the measures listed in 2.1.1.

2.1.14 Measures of entropy and their Properties :

In the following table we have measures of entropy listed in 1.1 in rows and their properties listed in 1.2 in columns.

The following symbols are made use of :

- | | | |
|------|---|---|
| Yes | - | if the property holds for the measure |
| NO | - | if the property does not hold for the measure |
| Yes/ | - | if the property holds conditionally or for a restricted range of the parameter(s) of the measure, |
| --- | - | if we are inclusive about the validity of the property for the measure. |
| X | - | does not apply. |

2.2 Measures of Directed Divergence

2.2.1 List of Measures of Directed Divergence

Let $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ be any two discrete probability distributions with $\sum_{i=1}^n p_i = 1$ and $0 \leq p_i < 1 \forall i = 1, \dots, n$. Then with the convention that $0 \log 0 = 0$, we have the following measures of directed divergence.

1) Kullback-Leiblers' Measure [45]

$$D(P;Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} \quad (2.1)$$

2) Renyi's Measure [55]

$$D_\alpha(P;Q) = \frac{1}{\alpha-1} \ln \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} \quad \alpha \neq 1 \quad (2.2)$$

3) Havrda-Charvats' Measure [8]

$$D^\alpha(P;Q) = \frac{1}{\alpha-1} \left[\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} - 1 \right] \quad \alpha \neq 1 \quad (2.3)$$

4) Sharma and Guptas' Measures [59]

$$(i) \quad D^L(P;Q;\alpha, \beta) = 2^{-\beta} \sum_{i=1}^n p_i^\alpha q_i^\beta \ln \frac{p_i}{q_i} \quad (2.4)$$

$$(ii) \quad D^P(P;Q;\alpha, \beta, \gamma) = (2^{\alpha-\beta-2\gamma-\beta})^{-1} \sum_{i=1}^n (p_i^\alpha q_i^{\beta-\alpha} - p_i^\gamma q_i^{\beta-\gamma}) \quad (2.5)$$

$$(iii) \quad D^S(P;Q;\beta, \gamma) = \frac{2^\beta}{\sin \gamma} \sum_{i=1}^n p_i q_i^\beta \sin \left(\gamma \ln \frac{p_i}{q_i} \right) \quad (2.6)$$

5) Ferrari's Measure [7]

$$D_{\lambda}(P;Q) = \frac{1}{\lambda} \sum_{i=1}^n (1 + \lambda p_i) \ln \frac{(1 + \lambda p_i)}{(1 + \lambda q_i)} \quad (2.7)$$

6) Kapur's Measure [18]

$$D_{\alpha,\beta}(P;Q) = \frac{1}{(\alpha-1)} \ln \left(\sum_{i=1}^n p_i^{\alpha+\beta-1} q_i^{1-\alpha} / \sum_{i=1}^n p_i^{\beta} \right) \quad (2.8)$$

7) Rathie's (n+1) parameter Measure [52]

$$D_{\alpha}^{\beta_1, \dots, \beta_m}(P;Q) = \frac{1}{(\alpha-1)} \ln \left\{ \sum_{i=1}^n p_i^{\alpha+\beta_1-1} q_i^{1-\alpha} / \sum_{i=1}^n p_i^{\beta_1} \right\} \quad (2.9)$$

8) Kapur's Measures [26,27,43]

$$(i) \quad D_{k1}(P;Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} - \frac{1}{a} \sum_{i=1}^n (1+ap_i) \ln \left(\frac{1+ap_i}{1+aq_i} \right), a \geq -1 \quad (2.10)$$

$$(ii) \quad D_G(P;Q) = \frac{1}{(\alpha-1)} \frac{1}{k} \frac{1}{n^b} \left(\sum_{i=1}^n p_i^{\alpha} \right)^a \left(\sum_{i=1}^n q_i^{1-\alpha} \right)^b \left(\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} \right)^c \\ \left(\left(\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} \right)^k - 1 \right) \quad (2.11)$$

$\alpha \neq k \neq 1$, a, b , and c are real numbers

$$(iii) \quad D_{\varphi}(P;Q) = \frac{1}{\alpha-1} \left\{ \varphi \left\{ \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} \right\} - (1) \right\} \quad (2.12)$$

where $\varphi(\cdot)$ is a twice differentiable convex function.

$$(iv) \quad D_{\Psi, \varphi}(P;Q) = \Psi \left\{ \sum_{i=1}^n q_i \varphi \left(\frac{p_i}{q_i} \right) \right\} \text{ where } \Psi(\cdot) \text{ is a twice} \\ \text{differentiable convex function} \quad (2.13)$$

2.2.2. List of Properties of Measures of Directed Divergence :

The following properties are checked for the directed divergence measures listed in 2.1. If $D(P;Q)$ is a directed divergence measure then,

- 1) $D(P;Q)$ is non-negative
- 2) $D(P;Q) = 0$ if and only if $P = Q$.
- 3) $D(P;Q)$ is a convex or pseudo-convex function of P and Q .
- 4) Additivity : If P and Q and R and S are pairwise independent probability distributions then

$$D(P * Q; R * S) \leq D(P;R) + D(Q;S). \text{ [Here } P * Q = PQ \text{ and } R * S = RS]$$
- 5) Subadditivity : If P, Q, R and S are any probability distributions then

$$D(P * Q; R * S) \leq D(P;R) + D(Q;S).$$
- 6) For $0 \leq \lambda \leq 1$ and P and Q any probability distributions
 $f(\lambda) = D[P; \lambda Q + (1-\lambda)P]$ is an increasing function of λ .
- 7) If any parameters are involved then $D(P;Q)$ is a monotonic function of each of them.
- 8) If any parameters are involved then $D(P;Q)$ is either concave (pseudo-concave) or convex (pseudo-convex) with respect to each of them.

Analysis

2.2.3 Kullback-Leibler Measure of Directed Divergence

$D(P;Q)$ satisfies all the properties listed in 2.2.2 except for the properties with respect to the parameters. In what

follows, we shall verify (i) Subadditivity and (ii) the monotonically increasingness of $f(\lambda) = D(P; \lambda Q + (1-\lambda)P)$. The latter is proved by Kapur [14].

Subadditivity of $D(P; Q)$: Let $P * Q$ and $R * S$ be any two joint probability distributions with P and Q and R and S as their respective marginal distributions. Then we have

$$D(P * Q; R * S) = \sum_{i=1}^n \sum_{j=1}^n p_i q_{ji} \ln \frac{p_i q_{ji}}{r_i s_{ji}}$$

where $(P * Q) = (\pi_{ij})_{n \times n} = (p_i q_{ji})_{n \times n}$ and $R * S = (\pi'_{ij})_{n \times n}$

$$= (r_i s_{ji})_{n \times n}$$

$$= \sum_{i=1}^n p_i \ln \frac{p_i}{r_i} + \sum_{i=1}^n p_i \sum_{j=1}^n q_{ji} \ln \frac{q_{ji}}{s_{ji}}$$

$$= \sum_{i=1}^n p_i \ln \frac{p_i}{r_i} + \sum_{i=1}^n p_i \sum_{j=1}^n q_{ji} \ln \frac{q_{ji}}{s_{ji}} \left[\sum_{j=1}^n q_{ji} = 1 \right]$$

$$= D(P; R) + \sum_{i=1}^n p_i D(Q; S/P_i)$$

$$\leq D(P; R) + D(Q; S).$$

That proves the subadditivity of $D(P; Q)$. Now we shall reproduce here the elegant proof of Kapur.

Let $f(\lambda) = D\{P; \lambda Q + (1-\lambda)P\}$

$$= \sum_{i=1}^n (\lambda q_i + (1-\lambda)p_i) \varphi\left(\frac{p_i}{\lambda q_i + (1-\lambda)p_i}\right) \text{ where } \varphi(x) = x \ln x.$$

We know that $\varphi(x)$ is a convex function and so $\varphi''(x) \geq 0$.

We have

$$f''(\lambda) = \sum_{i=1}^n \frac{p_i (q_i - p_i)^2}{(q_i + (1-\lambda)p_i)^2} \varphi''\left(-\frac{p_i}{q_i + (1-\lambda)p_i}\right) \geq 0.$$

Because $f'(\lambda) = \sum_{i=1}^n \frac{p_i (q_i - p_i)}{q_i + (1-\lambda)p_i}$ is zero when $\lambda = 0$, we get that $f'(\lambda) \geq 0$ which is the desired result.

With that we conclude our discussion of Kullback-Leibler measure of directed divergence.

2.2.4 Renyi's Measure of Directed Divergence : It is given by

$$D_\alpha(P;Q) = \frac{1}{\alpha-1} \ln \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} \quad \alpha \neq 1 \quad (2.14)$$

We can easily observe that $D_\alpha(P;Q)$ is non-negative by making use of Renyi's inequality [15], viz.

$$\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} \geq 1 \text{ according as } \alpha > 1, \quad \sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1 \quad (2.15)$$

$D_\alpha(P;Q)$ is pseudo-convex because $\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}$ is convex for $\alpha > 1$ and is concave for $0 < \alpha < 1$, and \log (convex function) is pseudo-convex function and \log (concave function) is pseudo-concave function. Now if $\alpha > 1$, $(\alpha-1)$ is positive and if $\alpha < 1$, $(\alpha-1)$ is negative. Again negative of a pseudo-concave function is pseudo-convex. So that makes, for all α , Renyi's measure of directed divergence is pseudo-convex.

Example 2.2.4.1 : Let $P * Q = \begin{pmatrix} 0.02 & 0.08 \\ 0.27 & 0.63 \end{pmatrix}$ and $R * S = \begin{pmatrix} 0.03 & 0.06 \\ 0.30 & 0.61 \end{pmatrix}$

(i) $\alpha = 2$. Then we have

$$D_2(P * Q; R * S) = 0.013563338, D_2(P; R) = 0.00122025511 \text{ and}$$

$$D_2(Q; R) = 0.0072104861$$

Therefore $D_2(P * Q; R * S) > D_2(P; R) + D_2(Q; S)$.

(ii) Let $\alpha = 0.5$. Then we have

$$D_{\frac{1}{2}}(P * Q; R * S) = 0.5056, D_{\frac{1}{2}}(P; R) = 0.1226 \text{ and } D_{\frac{1}{2}}(Q; S) = 0.3816$$

$$\text{again we have } D_{\frac{1}{2}}(P * Q; R * S) > D_{\frac{1}{2}}(P; R) + D_{\frac{1}{2}}(Q; S).$$

We have made calculations using various distributions for subadditivity of Renyi's measure of D.D. our results are tabulated low :

$$D_1 : P * Q = \begin{pmatrix} 0.2 & 0.3 \\ 0.2 & 0.3 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.1 & 0.6 \\ 0.1 & 0.2 \end{pmatrix}$$

α	$D_\alpha(P * Q; R * S) - D_\alpha(P; R) - D_\alpha(Q; S)$
0.110	0.2270182×10^{-2}
0.510	0.6353982×10^{-2}
0.960	0.2983892×10^{-3}
1.010	$-0.1178731 \times 10^{-2}$
1.510	$-0.2481915 \times 10^{-1}$
2.010	$-0.6186135 \times 10^{-1}$

Table 2.2.4.1

For D_1 Renyi's measure of directed divergence is not subadditive for $0 \leq \alpha < 1$ and is subadditive for all $\alpha \geq 1$

$$D_2 : P * Q = \begin{pmatrix} 0.15 & 0.35 \\ 0.25 & 0.25 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.05 & 0.65 \\ 0.15 & 0.15 \end{pmatrix}.$$

α	$D_\alpha(P * Q; R * S) - D_\alpha(P; R) - D_\alpha(Q; S)$
0.150	0.2988417×10^{-2}
0.550	0.9474083×10^{-2}
0.950	0.1181955×10^{-1}
2.135	0.1164209×10^{-3}
2.140	$-0.1256298 \times 10^{-3}$
2.175	$-0.1822637 \times 10^{-2}$

Table 2.2.4.2

For D_2 , Renyi's directed divergence of order α is not subadditive till α reaches a value α_0 , $2.135 < \alpha_0 < 2.140$. For $\alpha \geq \alpha_0$, it is subadditive.

$$D_3 : P * Q = \begin{pmatrix} 0.175 & 0.325 \\ 0.225 & 0.275 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.075 & 0.625 \\ 0.125 & 0.175 \end{pmatrix}$$

α	$D_\alpha(P * Q; R * S) - D_\alpha(P; R) - D_\alpha(Q; S)$
0.150	0.2556981×10^{-2}
0.550	0.5808485×10^{-2}
0.950	0.1620397×10^{-2}
1.015	0.9306082×10^{-4}
1.020	$-0.2774503 \times 10^{-4}$
1.025	$-0.1581162 \times 10^{-2}$
1.075	0.1524230×10^{-2}

Table 2.2.4.3

We observe that for D_3 Renyi's measure of directed divergence is not subadditive for all $\alpha < \alpha_0$ where $1.015 < \alpha_0 < 1.020$ and is subadditive for all $\alpha \geq \alpha_0$.

$$D_4 : P * Q = \begin{pmatrix} 0.02 & 0.08 \\ 0.27 & 0.63 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.03 & 0.06 \\ 0.30 & 0.61 \end{pmatrix}$$

Our results show that Renyi's measure of directed divergence is not subadditive for this set of distributions for any value of α .

$$D_5 : P * Q = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.02 & 0.08 \\ 0.83 & 0.07 \end{pmatrix}$$

α	$D_\alpha(P * Q; R * S) - D_\alpha(P; R) - D_\alpha(Q; S)$
0.150	0.2016853×10^{-1}
0.550	0.7489688×10^{-1}
0.950	0.9015147×10^{-1}
1.010	0.8628271×10^{-1}
1.510	0.7976070×10^{-3}
1.560	$-0.1083849 \times 10^{-1}$
1.760	$-0.5797683 \times 10^{-1}$

Table 2.2.4.4

For D_5 we find that Renyi's measure of directed divergence is not subadditive for $\alpha < \alpha_0$ where $1.510 < \alpha_0 < 1.560$ and is subadditive for all $\alpha \geq \alpha_0$.

Kapur [19] has established that $D_\alpha(P; Q)$ is a monotonically increasing function of α and also that it is a pseudo-convex

function α for $0 < \alpha < 1$ and a pseudo-concave function of α for $\alpha > 1$.

Now we shall conclude our discussion of Renyi's measure of directed divergence with our result that $D_\alpha(P * Q; R * S)$ satisfies property (6) only for $0 < \alpha < 1$.

Proposition 2.2.4.1 : $f(\lambda) = D_\alpha\{P; \lambda Q + (1-\lambda)P\}$, $0 \leq \lambda \leq 1$, is a monotonically increasing function of λ if $0 \leq \alpha < 1$.

Proof : We have from definition of $D_\alpha(P; Q)$,

$$f(\lambda) = \frac{1}{\alpha-1} \ln \sum_{i=1}^n p_i^\alpha \{\lambda q_i + (1-\lambda)p_i\}^{1-\alpha} \quad \text{and therefore}$$

$$f'(\lambda) = \frac{1}{\alpha-1} \frac{\sum_{i=1}^n (1-\alpha) p_i^\alpha \{\lambda q_i + (1-\lambda)p_i\}^{1-\alpha} (q_i - p_i)}{\sum_{i=1}^n p_i^\alpha \{\lambda q_i + (1-\lambda)p_i\}^{1-\alpha}} \quad \text{and} \quad (2.16)$$

$$\begin{aligned} & \{-\alpha(1-\alpha) \left(\sum_{i=1}^n p_i^\alpha \{\lambda q_i + (1-\lambda)p_i\}^{1-\alpha} \right) \left(\sum_{i=1}^n p_i^\alpha (\lambda q_i + (1-\lambda)p_i)^{-(\alpha+1)} \right) \\ & \times (q_i - p_i)^2 - (1-\alpha)^2 \left\{ \sum_{i=1}^n p_i^\alpha (\lambda q_i + (1-\lambda)p_i)^{-\alpha} (q_i - p_i) \right\}^2 \} \\ f''(\lambda) = & \frac{1}{\alpha-1} \frac{\sum_{i=1}^n p_i^\alpha (\lambda q_i + (1-\lambda)p_i)^{1-\alpha}}{\left\{ \sum_{i=1}^n p_i^\alpha (\lambda q_i + (1-\lambda)p_i)^{1-\alpha} \right\}^2} \end{aligned} \quad (2.17)$$

We can easily see that $f'(0) = 0$. For $0 \leq \alpha \leq 1$, we have from (2.17) that $f''(\lambda) \geq 0$. Therefore for $0 \leq \alpha \leq 1$, $f'(\lambda)$ is increasing from zero as λ increases from zero. That is $f'(\lambda) \geq 0$ for $0 \leq \lambda \leq 1$. But for $\alpha > 1$, we can not conclude any similar result.

2.2.5 Havrda-Charvat Measure of Directed Divergence :

It is given by

$$D^{\alpha}(P;Q) = \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} - 1 \right\} \quad \alpha \neq 1, \alpha > 0. \quad (2.18)$$

It can be easily verified that $D^{\alpha}(P;Q)$ is non-negative by making use of (2.16) and convex w.r. to both P and q by making use of the fact that $\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}$ is convex for $\alpha > 1$ and concave for $\alpha < 1$.

It can also be verified that it is a non-additive measure unlike $D_{\alpha}(P;Q)$ and $D(P;Q)$. We now consider its subadditivity property :

Example 2.2.5.1 : Let $P * Q = \begin{pmatrix} 0.02 & 0.08 \\ 0.27 & 0.63 \end{pmatrix}$ and $R * S = \begin{pmatrix} 0.03 & 0.06 \\ 0.30 & 0.61 \end{pmatrix}$

and $P = (0.1, 0.9)$, $Q = (0.29, 0.71)$, $R = (0.09, 0.91)$ and $S = (0.33, 0.67)$.

(i) $\alpha = 2$, then we have

$$D^2(P * Q; R * S) = 0.0136557, D^2(P; R) = 0.001221 \text{ and}$$

$$D^2(Q; R) = 0.0072365$$

$$\text{Therefore } D^2(P * Q; R * S) > D^2(P; R) + D^2(Q; S)$$

(ii) $\alpha = \frac{1}{2}$, then we have

$$D^2(P * Q; R * S) = 0.0033974, D^2(P; R) = 0.00029098 \text{ and}$$

$$D^2(Q; S) = 0.0018715.$$

$$\text{Again we have } D^2(P * Q; R * S) > D^2(P; R) + D^2(Q; S).$$

Table 2.2.5.1

$$D_1 : P * Q = \begin{pmatrix} 0.2 & 0.3 \\ 0.2 & 0.3 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.1 & 0.6 \\ 0.1 & 0.2 \end{pmatrix}$$

Then our calculations have shown that for all $\alpha < 1$, Havrda-Charvats' measure of directed divergence is not subadditive while for all $\alpha > 1$, it is subadditive.

Table 2.2.5.2

$$D_2 : P * Q = \begin{pmatrix} 0.15 & 0.35 \\ 0.25 & 0.25 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.05 & 0.65 \\ 0.15 & 0.15 \end{pmatrix}$$

Then for D_2 our calculations have shown that Havrda-Charvats' measure of directed divergence is not subadditive for any value of α .

Table 2.2.5.3 :

$$D_3 : P * Q = \begin{pmatrix} 0.175 & 0.325 \\ 0.225 & 0.275 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.075 & 0.625 \\ 0.125 & 0.175 \end{pmatrix}$$

α	$D^\alpha(P * Q; R * S) - D^\alpha(P; R) - D^\alpha(Q; S)$
0.100	0.1668872×10^{-2}
0.500	0.4463673×10^{-2}
0.900	0.1856760×10^{-2}
1.100	$-0.1177340 \times 10^{-2}$
2.000	$-0.2599998 \times 10^{-1}$
5.600	$-0.1707166 \times 10^{-2}$
5.700	0.2041808×10^{-1}
7.500	1.448590
9.900	16.828800

Table 2.2.5.2

We observe that for $\alpha < \alpha_0$ where $1.0 < \alpha_0 < 1.10$, D^α is not subadditive then for $\alpha_0 < \alpha < \alpha_1$ where $5.6 < \alpha_1 < 5.7$, D^α is subadditive and again for $\alpha > \alpha_1$, D^α is not subadditive.

Table 2.2.5.4 :

$$D_4 : P * Q = \begin{pmatrix} 0.02 & 0.08 \\ 0.27 & 0.63 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.03 & 0.06 \\ 0.30 & 0.61 \end{pmatrix}$$

Our calculations have shown that D^α is not subadditive for any value of α .

Table 2.2.5.5 :

$$D_5 : P * Q = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.02 & 0.08 \\ 0.83 & 0.07 \end{pmatrix}$$

For D_5 also our calculations have shown that D^α is not subadditive for any value of α .

Now we shall consider the properties of $D^\alpha(P;Q)$ with respect to α . Kapur [18] has established that $D^\alpha(P;Q)$ is a monotonically increasing with α and is a pseudo-convex function of α for $0 < \alpha < 1$ and pseudo-concave for $\alpha > 1$. We shall now consider the verification of Property (6) of 2.2.2.

Proposition 2.2.5.1 : $f(\lambda) = D^\alpha\{P; \lambda Q + (1-\lambda)P\}$ is an increasing function of λ for $0 \leq \lambda \leq 1$.

Proof : We have $f(\lambda) = \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i^\alpha (\lambda q_i + (1-\lambda)p_i)^{1-\alpha} - 1 \right\}$

$$f'(\lambda) = \frac{1}{\alpha-1} \sum_{i=1}^n p_i^\alpha (1-\alpha) (\lambda q_i + (1-\lambda)p_i)^{-\alpha} (q_i - p_i) \quad (2.20)$$

$$f''(\lambda) = \frac{-\alpha(1-\alpha)}{(\alpha-1)} \sum_{i=1}^n p_i^\alpha (\lambda q_i + (1-\lambda)p_i)^{-(\alpha+1)} (q_i - p_i)^2 \quad (2.21)$$

From (2.20) we can see that $f'(0) = 0$ and from (2.21) we get that $f''(\lambda) \geq 0$ for all values of $\alpha \geq 0$. Therefore we conclude that $f'(\lambda)$ increases from 0 as λ increases from 0 to 1.

With that we conclude our discussion of the Havrda-Charvat measure of directed divergence.

2.2.6 Sharma and Guptas' Measures of Directed Divergence :

These measures are given by

$$D^L(P;Q;\alpha,\beta) = \frac{1}{2^\beta} \sum_{i=1}^n p_i^\alpha q_i^\beta \ln \frac{p_i}{q_i} \quad (2.22)$$

$$D^P(P;Q;\alpha,\beta,\gamma) = \frac{1}{(2^{\alpha-\beta-2\gamma-\beta})} \sum_{i=1}^n (p_i^\alpha q_i^{\beta-\alpha} - p_i^\gamma q_i^{\beta-\gamma}) \quad (2.23)$$

$$D^S(P;Q;\beta,\gamma) = \frac{2^\beta}{\sin \gamma} \sum_{i=1}^n p_i^\gamma q_i^\beta \sin \left(\gamma \ln \frac{p_i}{q_i} \right) \quad (2.24)$$

Not all three measures are non-negative. Refer Kapur [24] for the following results :

(i) For $D^L(P;Q;\alpha,\beta)$ it can easily be verified that

$$2^{-\beta} D^L(P;Q;\alpha,\beta) + 2^{-\alpha} D^L(Q;P;\alpha,\beta) = 0$$

Therefore for every positive $D^L(P;Q;\alpha,\beta)$ we have a negative $D^L(Q;P;\alpha,\beta)$.

(ii) Let $\alpha = 1, \beta = 2, \gamma = 2$ then (2.23) gives

$$D^P(P;Q;1,2,2) = 2 \sum_{i=1}^n (p_i^2 - p_i q_i) \quad (2.25)$$

Now if we let $P = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ and $Q = (\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$ in (2.25)* we get

$$D^P(P:Q;1,2,2) = -0.083333 < 0.$$

Again if we take $P = U$ and Q arbitrary in (2.25) we get

$$D^P(U:Q;1,2,2) = \frac{2}{n} \left(1 - \sum_{i=1}^n q_i\right) = 0.$$

So $D^P(P:Q;\alpha,\beta,\gamma)$ can be negative and zero even if $P \neq Q$.

(iii) Let $\alpha = 1$, $\beta = 0$, $\gamma = 4$, $n = 2$, $P = (.5, .5)$ and $Q = (.4, .6)$ then from (2.24) we get

$$D^S(P:Q;1,0,4) = -0.07458 < 0$$

Again if we take $P = (.9, .1)$ and $Q = (.1, .9)$ we get

$D^S(P:Q;\alpha,\beta,\gamma) = 0$. Therefore $D^S(P:Q;\alpha,\beta,\gamma)$ can both be negative and vanish even if $P \neq Q$.

We can easily see that none of Sharma and Guptas' three measures is additive.

Now consider the following example :

Example 2.2.6.1 : Let $P * Q = \begin{pmatrix} 0.2 & 0.3 \\ 0.3 & 0.2 \end{pmatrix}$ and $R * S = \begin{pmatrix} 0.02 & 0.72 \\ 0.08 & 0.18 \end{pmatrix}$

$P = (0.5, 0.5)$, $Q = (0.4, 0.6)$, $R = (0.2, 0.8)$ and $S = (0.1, 0.9)$

and $\alpha = \beta = 2$. Then we have

$$D^L(P * Q : R * S; 2, 2) = -0.0099777,$$

$$D^L(P : R; 2, 2) = -0.0165904 \quad \text{and} \quad D^L(Q : S; 2, 2) = -0.0290038.$$

Therefore we have $D(P * Q : R * S; 2, 2) > D^L(P : R; 2, 2) + D^L(Q : S; 2, 2)$.

With the same example if we take $\gamma = 1$ we get a counter example for the subadditivity of $D^P(P:Q;\alpha,\beta,\gamma)$ and if we take $\beta = 2$, $\gamma = 2$ we get a counter-example for the subadditivity

of $D^S(P:Q;\beta,\gamma)$. So those examples assert that none of Sharma and Guptas' three measures of directed divergence is sub-additive.

We now present the results of our numerical calculations for various distributions regarding the subadditivity of Sharma Gupta log measure of directed divergence.

Table 2.2.6.1 :

α	β	$D^L(P * Q; R * S) - D^L(P; R) - D^L(Q; S).$
0.1	0.200	0.392063
	0.500	0.5046456×10^{-1}
	0.600	$-0.1652725 \times 10^{-1}$
	1.900	-0.1750458
0.3	0.200	0.2684716
	0.500	0.1977368×10^{-1}
	0.600	$-0.2879094 \times 10^{-1}$
	1.900	-0.1383785
0.6	0.200	0.1495838
	0.400	0.3409685×10^{-1}
	0.500	$-0.5014992 \times 10^{-2}$
	1.900	$-0.9717659 \times 10^{-1}$
1.0	0.200	0.6559744×10^{-1}
	0.400	0.4145582×10^{-1}
	0.500	$-0.1653739 \times 10^{-1}$
	1.900	$-0.6057082 \times 10^{-1}$
1.5	0.200	0.2062240×10^{-1}
	0.300	0.4823544×10^{-2}
	0.400	$-0.7387305 \times 10^{-2}$
	1.900	$-0.3348142 \times 10^{-1}$

In fact we have corresponding to every α lying between 0.1 and 1.9 with 0.1 increments, values of β where D^L changes from being not subadditive to being subadditive. Similarly we have values of α and β for data D_2, D_3, D_4 and D_5 . It is highly tedious to present all values here. Therefore we present all the important information in the following table :

Data	α	Critical interval for β	Data	α	Critical interval for β	Data	α
D_1	0.1	0.5-0.6	D_2	0.6	0.4-0.5	D_3	1.2	0.3-0.4
	0.2			0.7			1.3	
	0.3			0.8			1.4	0.2-0.3
	0.4			0.9	1.5			
	0.5			1.0	1.6			
	0.6	1.1		1.7				
	0.7	1.2		1.8				
	0.8	1.3		1.9				
	0.9	0.4-0.5		1.4	D_5	1.9	0.1-0.2	
	1.0			1.5		0.1		
	1.1			1.6		0.2		
	1.2			1.7		0.3		
	1.3			1.8		0.4		
	1.4	1.9		0.5				
1.5	0.3-0.4	0.1	0.6					
1.6		0.2	0.7					
1.7		0.3	0.8					
1.8		0.4	0.9					
1.9		0.5	1.0					
D_2	0.1	0.6-0.7	D_3	0.6	0.5-0.6	1.1	0.4-0.5	
	0.2			0.7		1.2		
	0.3			0.8		1.3		
	0.4			0.9	1.4			
	0.5			1.0	1.5			
		0.4-0.5		1.1	1.6			
				1.2	1.7			
				1.3	1.8			
				1.4	1.9			
				1.5				

Table 2.2.62

We shall now consider the properties of these measures with respect to their parameters.

Monotonous behaviour of $D^L(P;Q;\alpha,\beta,Y)$ as a function of α

Let $f(\alpha) = D^L(P;Q;\alpha,\beta,X)$. Then we have

$$f'(\alpha) = e^{-\beta} \sum_{i=1}^n p_i^\alpha q_i^\beta \{ \log(p_i)^2 - \log p_i \log q_i \} \quad (2.26)$$

We will now show with the help of an example that $f'(\alpha)$ can both be negative and positive.

Example 2.2.6.2 : Let $\alpha = 2$, $\beta = 2$, $P = (0.5, 0.5)$ and $Q = (0.6, 0.4)$. Then we have from (2.26)

$$f'(\alpha) = 0.00129674 > 0.$$

Again let $\alpha = 2$, $\beta = 2$, $P = (0.5, 0.5)$ and $Q = (0.55, 0.45)$. Then we have from (2.26)

$$f'(\alpha) = -8.8168 \times 10^{-5} < 0.$$

This example shows that $f'(\alpha)$ can be positive and negative depending on the distributions, for the same α . Therefore we deduce that $f(\alpha)$ does not exhibit a monotonic behaviour.

Monotonous behaviour of $D^L(P;Q;\alpha,\beta,Y)$ as a function of β :

Let $g(\beta) = D^L(P;Q;\alpha,\beta,X)$. Then we have

$$g'(\beta) = 2^{-\beta} \sum_{i=1}^n p_i^\alpha q_i^\beta \log\left(\frac{p_i}{q_i}\right) \left\{ \log q_i - \frac{\beta}{2} \right\} \quad (2.27)$$

Example 2.2.63 : Let $\alpha = 2$, $\beta = 2$, $P = (0.5, 0.5)$ and $Q = (0.55, 0.45)$ in (2.27). Then we have $g'(\beta) = 4.8097 \times 10^{-4} > 0$.

Again let $\alpha = 2$, $\beta = 2$, $P = (0.5, 0.5)$ and $Q = (0.6, 0.4)$ in (2.27). Then we have $g'(\beta) = -0.4833487 < 0$.

This example illustrates that $g(\beta)$ is not a monotonic function of β .

Monotonous behaviour of $D^P(P; Q; \alpha, \beta, \gamma)$ as a function of α :

Let $f_1(\alpha) = D^P(P; Q; \alpha, \beta, \gamma)$ then we have

$$\begin{aligned} f_1'(\alpha) = & (2^{\alpha-\beta} 2^{\gamma-\beta})^{-1} \sum_{i=1}^n (p_i^\alpha q_i^{\beta-\alpha} \ln \frac{p_i}{q_i}) \\ & - (2^{\alpha-\beta} 2^{\gamma-\beta})^{-2} \ln 2 \cdot 2^{\alpha-\beta} \sum_{i=1}^n (p_i^\alpha q_i^{\beta-\alpha} - p_i^\gamma q_i^{\beta-\gamma}) \end{aligned} \quad (2.28)$$

Example 2.2.6.4 : Let $\alpha = 1.5$, $\beta = 0.5$, $\gamma = 0$, $P = (0.5, 0.5)$ and $Q = (0.55, 0.45)$ in (2.28).

Then we have $f_1'(\alpha) = 0.2599682 > 0$.

Again let $\alpha = 1.5$, $\beta = 0.5$, $\gamma = 0$, $P = (0.2, 0.8)$ and $Q = (0.5, 0.5)$ in (2.28). Then we have $f_1'(\alpha) = -1.5666427 < 0$.

This example shows that $f(\alpha)$ is not monotonic with respect to α .

Monotonous behaviour of $D^P(P; Q; \alpha, \beta, \gamma)$ as a function of β :

Let $g_1(\beta) = D^P(P; Q; \alpha, \beta, \gamma)$. Then we have

$$\begin{aligned} g_1'(\beta) = & (2^{\alpha-\beta} 2^{\gamma-\beta})^{-1} \sum_{i=1}^n (p_i^\alpha q_i^{\beta-\gamma} - p_i^\gamma q_i^{\beta-\gamma}) \ln q_i \\ & + (2^{\alpha-\beta} 2^{\gamma-\beta})^{-2} (2^{\alpha-\beta} + 2^{\gamma-\beta}) \ln 2 \sum_{i=1}^n (p_i^\alpha q_i^{\beta-\alpha} - p_i^\gamma q_i^{\beta-\gamma}) \end{aligned} \quad (2.29)$$

Example 2.2.6.5 : Let $\alpha = 1.5$, $\beta = 0.5$, $\gamma = 0$, $P = (0.2, 0.8)$ and $Q = (0.5, 0.5)$. Then from (2.29) we have $g'_1(\beta) = 0.1147965 > 0$. Again let $\alpha = 1.5$, $\beta = 0.5$, $\gamma = 0$, $P = (0.5, 0.5)$ and $Q = (0.6, 0.4)$. Then from (2.29) we have $g'_1(\beta) = -0.0307757 < 0$.

Again we have that $D^P(P; Q; \alpha, \beta, \gamma)$ is not monotonic w.r. to β . Note that $D^S(P; Q; \alpha, \beta, \gamma)$ is symmetric in α and γ . Therefore our conclusions regarding the monotonicity behaviour of it w.r. to α hold good for that w.r. to γ also. Therefore we conclude that Sharma and Guptas' power measure of directed divergence is not a monotonic function of any of its three parameters.

Monotonous behaviour of $D^S(P; Q; \beta, \gamma)$ as a function of β :

Let $f_2(\beta) = D^S(P; Q; \beta, \gamma)$. Then we have

$$f'_2(\beta) = 2^\beta (\sin \gamma)^{-1} \sum_{i=1}^n p_i q_i^\beta \sin \left(\gamma \ln \frac{p_i}{q_i} \right) \ln (2q_i). \quad (2.29)$$

Example 2.2.6.6 : Let $\beta = 1.5$, $\gamma = 2$, $P = (0.5, 0.5)$ and $Q = (0.55, 0.45)$. From (2.29) we get $f'_2(\beta) = -0.0109112 < 0$. Again let $\beta = 1.5$, $\gamma = 2$, $P = (0.2, 0.8)$ and $Q = (0.9, 0.1)$. From (2.29) we get $f'_2(\beta) = 0.778836 > 0$.

Monotonous behaviour of $D^S(P; Q; \beta, \gamma)$ as a function of γ :

Let $g_2(\gamma) = D^S(P; Q; \beta, \gamma)$. Then we get

$$g'_2(\gamma) = 2^\beta (\sin \gamma)^{-1} \sum_{i=1}^n p_i q_i^\beta \{ \ln p_i \sin \left(\gamma \ln \frac{p_i}{q_i} \right) + \cos \left(\gamma \ln \frac{p_i}{q_i} \right) + \frac{\cos \gamma}{\sin \gamma} \sin \left(\gamma \ln \frac{p_i}{q_i} \right) \} \quad (2.30)$$

Example 2.2.6.7 : Let $\gamma = 1$, $\beta = 0.5$, $P = (0.5, 0.5)$ and $Q = (0.45, 0.55)$. Then from (2.30) we get $g_2'(\gamma) = -0.095183 < 0$.

Again let $\gamma = 1$, $\beta = 0.5$, $P = (0.2, 0.8)$ and $Q = (0.9, 0.1)$. Then from (2.30) we get $g_2'(\gamma) = 0.3666375 > 0$.

Therefore we conclude that $D^S(P; Q; \beta, \gamma)$ is not monotonic with respect to either of its parameters.

Concavity (or convexity) of $D^L(P; Q; \alpha, \beta)$ as a function of α :

$$\text{From (2.26), } f''(\alpha) = 2^{-\beta} \sum_{i=1}^n p_i^\alpha q_i^\beta (\log p_i)^2 \log\left(\frac{p_i}{q_i}\right) \quad (2.31)$$

Example 2.2.6.8 : Let $\alpha = 1$, $\beta = 0.2$, $P = (0.5, 0.5)$ and $Q = (0.99, 0.01)$. Then from (2.31) we get $f''(\alpha) = 0.1831302 > 0$.
Again let $\alpha = 1$, $\beta = 0.2$, $P = (0.85, 0.15)$ and $Q = (0.5, 0.5)$. Then from (2.31) we get $f''(\alpha) = -0.6287022 < 0$.

So we have seen from the example that for the same set of parametric values we have two sets of probability distribution such that $f''(\alpha)$ is positive for one and negative for the other. Therefore we conclude that $D^L(P; Q; \alpha, \beta)$ is neither concave nor convex for all α .

We have similar examples to show that $D^L(P; Q; \alpha, \beta)$ is not convex or concave w.r. to β .

We are unable to conclude anything about concavity (or convexity) of the other two measures, viz. $D^P(P; Q; \alpha, \beta, \gamma)$ and $D^S(P; Q; \beta, \gamma)$ because of the highly complicated nature of their second derivatives with respect to their parameters.

Similarly we are not able to decide whether or not $D^L(P:Q; \alpha, \beta)$ and $D^S(P:Q; \beta, \alpha)$ are convex w.r. to P and Q : whether or not they satisfy property no. (6).

But for $0 \leq \alpha \leq 1, \beta \geq 1$ or $\alpha \geq 1, 0 \leq \beta \leq 1, \alpha \neq \beta$. We have $D^P(P:Q; \alpha, \beta, \lambda)$ satisfying both convexity criterion and property (6). See Kapur [14].

With that we conclude our discussion of Sharma and Guptas' measures of directed divergence.

2.2.7 Ferrari's Measure of Directed Divergence :

It is given by

$$D(P:Q) = \frac{1}{\lambda} \sum_{i=1}^n (1 + \lambda p_i) \ln \left(\frac{1 + \lambda p_i}{1 + \lambda q_i} \right), \lambda > 0 \quad (2.32)$$

It can be easily verified that $D(P:Q)$ is a non-negative measure of directed divergence. It can also be verified that $D(P:Q)$ is a convex function of both P and Q , by considering the functions $f(x) = (1 + \lambda x) \{ \ln(1 + \lambda x) - A \}$ and $g(y) = B \{ \ln B - \ln(1 + \lambda y) \}$ and the fact that a finite sum of convex functions is convex.

We can easily verify that $D(P:Q)$ is not additive. We shall now give an example to show that $D(P:Q)$ is not sub-additive.

Example 2.2.7.1 : Let $\lambda = 1, P = (0.5, 0.5), Q = (0.5, 0.5), R = (0.1, 0.9)$ and $S = (0.85, 0.15)$ and let

$$P * Q = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix} \text{ and } R * S = \begin{pmatrix} 0.02 & 0.08 \\ 0.83 & 0.07 \end{pmatrix}.$$

Then we have $D_1(P * Q : R * S) = 0.1547$, $D_1(P : Q) = 0.0066$ and $D_1(Q : S) = 0.0269$.

Therefore $D_1(P * Q : R * S) > D_1(P : R) + D_1(Q : S)$.

Proposition 2.2.7.2 : $f(\lambda) = D_\lambda(P : Q)$ is monotonically decreasing for $0 \leq \lambda \leq \lambda_0$ and monotonically increasing for $\lambda \geq \lambda_0$ where λ_0 is a positive solution of the equation :

$$\sum_{i=1}^n \frac{p_i - q_i}{1 + \lambda q_i} - \frac{1}{\lambda} \sum_{i=1}^n \ln \frac{1 + \lambda p_i}{1 + \lambda q_i} = 0. \quad (2.33)$$

Proof : From the definition of $D_\lambda(P : Q)$ we have

$$f'(\lambda) = \frac{1}{\lambda} \left\{ \sum_{i=1}^n \left(\frac{p_i - q_i}{1 + \lambda q_i} + p_i \ln \frac{1 + \lambda p_i}{1 + \lambda q_i} \right) \right\} - \frac{1}{\lambda} f(\lambda) \quad (2.34)$$

We can verify easily by applying L'Hospital's rule that

$$\lim_{\lambda \rightarrow 0} f'(\lambda) = -\frac{1}{2} \sum_{i=1}^n (p_i - q_i)^2 < 0 \quad (2.35)$$

Now we shall consider the zeros of $f'(\lambda)$ for $\lambda > 0$. From (2.34) we get

$$\begin{aligned} f'(\lambda) = 0 & \text{ iff } \sum_{i=1}^n \left(\frac{p_i - q_i}{1 + \lambda q_i} + p_i \ln \frac{1 + \lambda p_i}{1 + \lambda q_i} \right) \\ &= \frac{1}{\lambda} \sum_{i=1}^n (1 + \lambda p_i) \ln \left(\frac{1 + \lambda p_i}{1 + \lambda q_i} \right) \end{aligned}$$

which is equivalent to (2.33).

We note that Equation (2.33) can have more than one solutions all of which are positive. In that case the sign of $f'(\lambda)$ keeps oscillating between plus and minus as we go from one zero to the next one. That completes the proof of

Proposition 2.2.7.2.

$D_\lambda(P:Q)$ satisfies the property no. (6) of 2.2.2, See [14].
With that we conclude our discussion of $D_\lambda(P:Q)$.

2.2.8 Kapur's Measure of Directed Divergence of Order α and Type β .

It is given by

$$D_{\alpha,\beta}(P:Q) = \frac{1}{\alpha-1} \ln \left\{ \sum_{i=1}^n p_i^{\alpha+\beta-1} q_i^{1-\alpha} / \sum_{i=1}^n p_i^\beta \right\} \quad \alpha \neq 1 \quad (2.36)$$

This measure corresponds to Kapur-Aczel and Daroczy measure of entropy of order α and type β , $H_{\alpha,\beta}(P)$.

Kapur et al. [15] have discussed the properties of this measure and shown that for $0 < \alpha < 1, \beta > 1$ or $0 < \beta < 1, \alpha > 1$ $\left. \begin{array}{l} 1 < \alpha + \beta < 2 \\ \alpha + \beta > 2 \end{array} \right\} D_{\alpha,\beta}(P:Q)$ is non-negative, vanishes only for $P = Q$ and is convex w.r. to both P and Q . We could not conclude anything regarding Property (6) for this measure.

Rathie's measure with $(n+1)$ parameters $D_{\alpha}^{\beta_1, \dots, \beta_n}(P:Q)$ is essentially same as (2.36) but for the n parameters β_i which were all equal in (2.36). Therefore if

$$\begin{array}{ll} 0 < \alpha < 1, \beta_i > 1 & \forall i = 1, 2, \dots, n \text{ or } 0 < \beta_i < 1, \alpha > 1 \\ 1 < \alpha + \beta_i < 2 & \alpha + \beta_i > 2 \end{array}$$

$$\forall i = 1, \dots, n$$

then $D_{\alpha}^{\beta_1, \dots, \beta_n}(P:Q)$ is non-negative zero, only if $P = Q$ and convex w.r. to both P and Q . But the major drawbacks of this

measure are that, it is not symmetric w.r. to the probabilities, and we can not define additivity and subadditivity properties for it.

2.2.9 Kapur's Generalized Measures of Directed Divergence :

$$(i) \quad D_{k1}(P:Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} - \frac{1}{a} \sum_{i=1}^n (1+ap_i) \ln \left(\frac{1+ap_i}{1+aq_i} \right) \quad a \geq -1 \quad (2.37)$$

$D_{k1}(P:Q)$ is a non-negative and pseudo-convex function of both P and Q . It vanishes iff $P = Q$. We can see easily that as a increases 0 to ∞ $D_{k1}(P:Q)$ decreases from $\sum_{i=1}^n p_i \ln \frac{p_i}{q_i}$ to zero. It does not satisfy property (6). It is neither additive nor subadditive. Consider the following example :

Example 2.2.9.1 : Let $P = (0.1, 0.9)$, $Q = (0.3, 0.7)$, $R = (0.45, 0.55)$ and $S = (0.5, 0.5)$. Let $a = 1$ and $P * Q = PQ$ and $R * S = RS$. Then we have

$$D_{k1}(P * Q : R * S) = 0.303043, \quad D_{k1}(P:R) = 0.209860 \text{ and}$$

$$D_{k1}(Q:S) = 0.055535.$$

$D_{k1}(P * Q : R * S) > D_{k1}(P:R) + D_{k1}(Q:S)$ which verified the claim that $D_{k1}(P:Q)$ is not a subadditive measure of directed divergence.

The remaining measures of directed divergence due to Kapur are very generalized measures, constructed to satisfy Properties (i), (ii) and (iii). Their significance lies in the fact that they contain many known measures as special cases.

Therefore here we only illustrate some special cases of each of these measures and refer the reader to Kapur [26] for further details on their properties :

(ii) $D_G(P:Q)$:

$$\text{Case 1 : } \lim_{\alpha \rightarrow 1} D_G(P:Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} = D(P:Q).$$

$$\text{Case 2 : } a = 0, b = 0, c = 0, k = 1, D_G(P:Q)$$

$$= \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} - 1 \right\} = D^\alpha(P:Q)$$

$$\text{Case 3 : } a = 0, b = 0, c = 0, k \rightarrow 0 D_G(P:Q) \rightarrow \frac{1}{\alpha-1} \ln \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}$$

$$= D_\alpha(P:Q)$$

$$\text{Case 4 : } a = 1, b = 0, c = 0, k = 1, D_G(P:Q)$$

$$= \frac{1}{(\alpha-1)} \left(\sum_{i=1}^n p_i^\alpha \right) \left(1 - \frac{1}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}} \right).$$

This is the error function obtained by Lubbe [22].

(iii) $D_\psi(P:Q)$:

$$\text{Case 1 : } D_{\alpha,j}(P:Q) = \frac{1}{\alpha-1} \left\{ \left(\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} \right)^j - 1 \right\}, \alpha > 1, j \geq 1 \text{ or } \alpha < 1, 0 < j \leq 1$$

For $j = 1$, $D_{\alpha,j}(P:Q)$ reduces to Havrda-Charvat measure.

$$\text{Case 2 : } D_\alpha(P:Q) = \frac{1}{\alpha-1} \ln \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}, \alpha \neq 1.$$

This is the well known Renyi's measure of directed divergence.

$$\text{Case 3 : } D_{\alpha, \theta, j, j'}(P:Q) = \frac{1}{\alpha - \beta} \left\{ \left(\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} \right)^j - \left(\sum_{i=1}^n p_i^\beta q_i^{1-\beta} \right)^{j'} \right\}$$

where $\alpha > 1, j \geq 1, \beta < 1, 0 \leq j' \leq 1$ or $\alpha < 1, 0 \leq j \leq 1,$

$\beta > 1, j' > 1$. If $j = j' = 1$ this measure reduces to Sharma and Tanejas' measure of directed divergence.

$$\text{Case 4 : } D_{\alpha, \beta}(P:Q) = \frac{1}{\alpha - \beta} \left\{ \frac{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}}{\sum_{i=1}^n p_i^\beta q_i^{1-\beta}} - 1 \right\}, \alpha > 1, \beta < 1$$

or $\alpha < 1, \beta > 1, \alpha \neq \beta$.

(iv) $D_{\Psi, \varphi}(P:Q)$:

Case 1 : If we take $\Psi(x) = \frac{x^\alpha - 1}{\alpha - 1}$, we get all the measures obtained as special cases for the measure $D_\varphi(P:Q)$ as special cases here also see [26].

Case 2 : If we take $\varphi(x) = x \ln x$, we get the limiting cases of special cases of $D_\varphi(P:Q)$ as $\alpha \rightarrow 1$, as special cases, here [26].

Case 3 : If we put $\Psi(x) = x \ln x - \frac{1}{a}(1+ax)\ln(1+ax) + \frac{1}{a}(1+a) \ln(1+a)$ then we get a function of Kapur's measure of directed divergence [17,26].

From these generalized measures of directed divergence, Kapur has also obtained measures of entropy by making use of

$$D(P:U) = H(U) - H(P)$$

where $U = (\frac{1}{n}, \dots, \frac{1}{n})$ the uniform distribution. Now we conclude this section by presenting a table, measures of directed

divergence vs their properties.

2.2.10 Measures of Directed Divergence and Their Properties

Properties Measures		1	2	3	4	5	6	7	8
1		Yes	Yes	Yes	Yes	Yes	Yes	X	X
2		Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
3		Yes	Yes	Yes	No	No	Yes	Yes	Yes
4	(i)	No	No	-	No	No	-	No	No
	(ii)	No	No	-	No	No	-	No	-
	(iii)	No	No	-	No	No	-	No	-
5		Yes	Yes	Yes	No	No	-	Yes/No	-
6		Yes	Yes	Yes/No	Yes	No	-	Yes/No	Yes/No
7		Yes	Yes	Yes/No	Yes	X	X	Yes/No	Yes/No
8	(i)	Yes	Yes	Yes	No	No	No	Yes/No	Yes/No
	(ii)	Yes	Yes	-	No	No	-	Yes/No	Yes/No
	(iii)	Yes	Yes	-	No	No	-	Yes/No	Yes/No
	(iv)	Yes	Yes	-	No	No	-	Yes/No	Yes/No

Table No. 2.2.10

The following convention is followed in Table No. 2.2.10.1

Yes - the property is satisfied

No - the property is satisfied

Yes/No - the property is conditionally satisfied

- - unknown.

X - Not applicable.

2.3 Measures of Inaccuracy

2.3.1 List of Measures of Inaccuracy

(1) Kerridge [44]

$$I(P:Q) = - \sum_{i=1}^n p_i \log q_i$$

(2) Nath [48]

$$I_{\alpha}(P:Q) = \frac{1}{1-\alpha} \log \sum_{i=1}^n p_i q_i^{1-\alpha}, \quad \alpha \neq 1, \alpha > 0$$

(3) Rathie and Kannappan [17]

$$I^{\alpha}(P:Q) = \frac{1}{1-\alpha} \left\{ \sum_{i=1}^n p_i q_i^{\alpha-1} - 1 \right\}, \quad \alpha \neq 1, \alpha > 0$$

Sharma and Gupta [59]

(4) Log measure

$$I_{\alpha, \beta}^L(P:Q) = - 2^{\beta} \sum_{i=1}^n p_i^{\alpha} q_i^{\beta} \log q_i, \quad \alpha > 0, \beta \geq 0$$

(5) Power measure

$$I_{\alpha, \beta}^P(P:Q) = \frac{1}{2^{-\beta-2}-\gamma} \sum_{i=1}^n p_i^{\alpha} (q_i^{\beta} - q_i^{\gamma}), \quad \alpha > 0, \beta, \gamma \geq 0, \beta \neq \gamma.$$

(6) Sine measure

$$I_{\alpha, \beta}^S(P:Q) = - \frac{2^{\beta}}{\sin \gamma} \sum_{i=1}^n p_i^{\alpha} q_i^{\beta} \sin(\log q_i), \quad \alpha > 0, \beta \geq 0, \gamma \neq 0$$

Kapur's measures [44, 60]

$$(7) \quad I_k(P:Q) = ((\alpha-1)k)^{-1} \left\{ n^{-b} \left(\sum_{i=1}^n p_i^{\alpha} \right)^a \left(\sum_{i=1}^n q_i^{1-\alpha} \right)^b \left(\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} \right)^c \right. \\ \left. \left(\left(\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} \right)^k - 1 \right) \right\} + n^{b+c+(\alpha-1)k} \{ 1 - n^{-k(\alpha-1)} \\ - \left(\sum_{i=1}^n p_i^{\alpha} \right)^{a+c+k} + \left(\sum_{i=1}^n p_i^{\alpha} \right)^{a+c} n^{-k(\alpha-1)} \}.$$

(8) If $a = b = c = 0, k = 1$ in (7), we get

$$I_{\alpha k}(P:Q) = \frac{1}{\alpha-1} \left(\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} - 1 \right) + \frac{n^{\alpha-1}}{1-\alpha} \left(\sum_{i=1}^n p_i^\alpha - 1 \right)$$

(9) If $a = b = c = 0, k \rightarrow 0$, in (7) we get Lubbe's [60] measure

$$I^\alpha(P:Q) = \frac{1}{\alpha-1} \log \left\{ \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} / \sum_{i=1}^n p_i^\alpha \right\}.$$

(10) If $a = b = c = 0, k = -1$, we get a new measure

$$I_N(P:Q) = \frac{1}{\alpha-1} \left(1 - \frac{1}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}} \right) + \frac{n^{1-\alpha}}{1-\alpha} \left(1 - \frac{1}{\sum_{i=1}^n p_i^\alpha} \right).$$

(11) If $a = 1, b = c = 0, k = -1$, we get Lubbe's [60] measure

$$I_L(P:Q) = \frac{1}{\alpha-1} \left(1 - \frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}} \right)$$

2.3.2 Properties of Measures of Inaccuracy :

Let $P = (p_1, p_2, \dots, p_n)$ be the true probability distribution and let $Q = (q_1, q_2, \dots, q_n)$ be the asserted probability distribution. Let $I(P:Q)$ be an inaccuracy measure. Then the following properties are verified for each of the measures listed in 2.3.1.

- I. The function I is continuous in p_i and q_i for all i .
- II. When N equally likely outcomes are stated to be equally likely then I is a monotonic increasing function of N .

- III. If a statement is broken down into a number of subsidiary statements the inaccuracy of the original statement is a weighted sum of the inaccuracies of the subsidiary statements. For example, we should have
- $$I(\alpha, \beta, \gamma; \theta, \varphi, \Psi) = I(\alpha, 1-\alpha; \theta, 1-\theta) + (1-\alpha)I\left(\frac{\beta}{1-\alpha}, \frac{\gamma}{1-\alpha}; \frac{\varphi}{1-\theta}, \frac{\Psi}{1-\theta}\right).$$
- IV. The inaccuracy of a statement is unchanged if two alternatives about which the same assertion is made are combined. For example,
- $$I(\alpha, \beta, \gamma; \theta, \varphi, \varphi) = I(\alpha, \beta + \gamma; \theta, \varphi).$$
- V. The quantity I is zero iff $p_i = q_i = 1$ for some value of i and consequently, $p_i = q_i = 0$ for all other values of i .
- VI. $I(P:Q)$ approaches infinity if $q_i = 0$ and the corresponding $p_i \neq 0$.
- VII. The value of I is minimum for a fixed $\{p_i\}$ when $q_i = p_i$ for all i . This value of inaccuracy is the amount of uncertainty involved in the distribution $\{p_i\}$.
- VIII. If variations of both p_i and q_i are considered the point $p_i = q_i = \frac{1}{n}$ is a minimax point.
- IX. If two sets of alternatives are asserted to have probabilities which are independent, the inaccuracy of the point assertion is the sum of the separate inaccuracies.
- X. Subadditivity : $I(P * Q : R * S) \leq I(P : R) + I(Q : S)$.
- XI. We require that $I(P:Q)$ is a convex function of Q .

2.3.3 Renyi's Measure of Inaccuracy :

P. Nath [48] introduced this measure. But due to its resemblance in form to Renyi's measure of information we call it Renyi's measure of inaccuracy. Let

$P = (p_1, p_2, \dots, p_n)$, $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$ be the true probability distribution of an expt. and let $Q = (q_1, \dots, q_n)$, $q_i \geq 0$, $\sum_{i=1}^n q_i = 1$ be the asserted probability distribution of the expt. Then Renyi's Inaccuracy is defined as

$$I_{\alpha}(P:Q) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n p_i q_i^{\alpha-1} \right) \quad (2.38)$$

is a continuous function of p_i 's and q_i 's for all i . We shall now verify the property II in the following :

Proposition 2.3.3.1 : When n equally likely outcomes are asserted to be equally likely then $I_{\alpha}(P:Q)$ is a monotonically increasing function of n , $n > 1$.

Proof : Let $P = Q = U = (\frac{1}{n}, \dots, \frac{1}{n})$. Then we have

$$\begin{aligned} I_{\alpha}(U:U) &= \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n \frac{1}{n} \left(\frac{1}{n} \right)^{\alpha-1} \right) \\ &= \ln n \end{aligned}$$

and we know that $\ln n$ is a monotonically increasing function for $n > 1$.

$I_{\alpha}(P;Q)$ does not have a recursive property. So now we verify the property IV.

Proposition 2.3.3.2 : $I_{\alpha}(p_1, p_2, p_3; q_1, q_2, q_2) = I_{\alpha}(p_1, p_2 + p_3; q_1, q_2)$

$$\begin{aligned}
\text{Proof : } I_{\alpha}(p_1, p_2, p_3; q_1, q_2, q_2) &= \frac{1}{1-\alpha} \ln(p_1 q_1^{\alpha-1} + p_2 q_2^{\alpha-1} + p_3 q_2^{\alpha-1}) \\
&= \frac{1}{1-\alpha} \ln(p_1 q_1^{\alpha-1} + (p_2 + p_3) q_2^{\alpha-1}) \\
&= I_{\alpha}(p_1, p_2 + p_3; q_1, q_2).
\end{aligned}$$

Proposition 2.3.3.3 : I_{α} is zero if and only if $p_i = q_i = 1$ for some value of i and $p_i = q_i = 0$ for all other values.

Proof : Sufficiency is obvious. We shall now prove the necessity part.

$$\text{Let } \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n p_i q_i^{\alpha-1} \right) = 0$$

$$\text{That means } \sum_{i=1}^n p_i q_i^{\alpha-1} = 1$$

$$\text{or } \sum_{i=1}^n p_i (q_i^{\alpha-1} - 1) = 0 \quad (2.39)$$

Now there are two possibilities. One is both P and Q are degenerate distributions with matching non-zero components and the other is that P and Q are any two distributions. Let us now consider the second possibility.

Then we interpret (2.39) as the average of n numbers $\{q_i^{\alpha-1} - 1\}_{i=1}^n$ which are all of the same sign (depending on $\alpha > 1$ or < 1 , they are negative or positive) with atleast two being non-zero. Hence the average cannot be zero. Therefore only the first possibility remains valid.

i.e., $p_i = q_i = 1$ for some i and $p_i = q_i = 0$ for all other i .

Property VI is obviously not satisfied. For consider

Example 2.3.3.1 : $P = (\frac{1}{2}, 0, \frac{1}{2})$ and $Q = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

Then we have $I_\alpha(P:Q) = \ln 4$.

Remark : In the definition of $I_\alpha(P:Q)$ Nath [48] had overlooked the point that negative exponent can not be raised to power zero. That is his definition included all $\alpha > 0$ except $\alpha = 1$. Then we can not have any q_i to be zero with a meaningful definition for $I_\alpha(P:Q)$. Hence what should be considered as valid range for the parameter values of α is $\alpha > 1$.

Property VII is satisfied. Let $p_i = q_i \forall i = 1, \dots, n$. Then we have $I_\alpha(P:P) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n p_i^\alpha \right)$, which is equal to the uncertainty involved in $P = \{p_i\}_1^n$. But this is not the minimum value of $I_\alpha(P:Q)$. We have obtained the minimum value of $I_\alpha(P:Q)$ in the following

Proposition 2.3.3.4 : The minimum value of $I_\alpha(P:Q)$ for a fixed P is attained when

$$q_i = \frac{p_i^{\frac{1}{2-\alpha}}}{\sum_{i=1}^n p_i^{1/2-\alpha}},$$

$$\min_Q \{I_\alpha(P:Q)\} = \frac{\alpha}{\alpha-1} \ln \left(\sum_{i=1}^n p_i^{\frac{1}{2-\alpha}} \right). \text{ As } \alpha \rightarrow 1,$$

$$\min_Q \{I_\alpha(P:Q)\} \rightarrow - \sum_{i=1}^n p_i \ln p_i$$

which is the minimum value of Kerridge's inaccuracy for a fixed P .

Proof : We shall note later that $I_\alpha(P:Q)$ is a convex function of q_i 's. Using that fact and Lagrange multipliers method we

shall obtain the minimum value of $I_\alpha(P:Q)$.

$$\text{Let } L \equiv \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n p_i q_i^{\alpha-1} \right) - \left(\sum_{i=1}^n p_i - 1 \right)$$

On equating $\frac{\partial L}{\partial q_i}$ to zero and solving for q_i we get

$$q_i = A p_i^{\frac{1}{2-\alpha}} \quad (2.40)$$

where the constant A is to be obtained from $\sum_{i=1}^n p_i = 1$. We obtain

$$q_i = \frac{p_i^{1/2-\alpha}}{\sum_{i=1}^n p_i^{1/2-\alpha}} \quad (2.41)$$

as p.d. which minimizes $I_\alpha(P:Q)$ for a fixed P . And

$$\text{Min}_Q \{I_\alpha(P:Q)\} = \frac{\alpha}{\alpha-1} \ln \left(\sum_{i=1}^n p_i^{\frac{1}{2-\alpha}} \right) \quad (2.42)$$

is obtained by substituting (2.41) in (2.38).

We now want to find $\lim_{\alpha \rightarrow 1} \text{Min}_Q \{I_\alpha(P:Q)\}$

$$= \lim_{\alpha \rightarrow 1} \frac{\alpha}{\alpha-1} \ln \left(\sum_{i=1}^n p_i^{\frac{1}{2-\alpha}} \right).$$

$$\begin{aligned} \text{Using L'Hospital's rule, } &= \lim_{\alpha \rightarrow 1} \left\{ -\alpha \frac{\sum_{i=1}^n p_i^{1/2-\alpha} \ln p_i}{\sum_{i=1}^n p_i^{1/2-\alpha}} + \ln \sum_{i=1}^n p_i^{\frac{1}{2-\alpha}} \right\} \\ &= -\sum_{i=1}^n p_i \ln p_i. \end{aligned}$$

Property VII is obviously not satisfied by (2.38) because it can not be expressed as a sum of entropy and directed

divergence, besides the uncertainty of the distribution P is not the minimum of $I_\alpha(P:Q)$. Hence there is no question of having a minimax point for $I_\alpha(P:Q)$.

$I_\alpha(P:Q)$ is an additive measure of inaccuracy. In otherwords $I_\alpha(P:Q)$ satisfies the property IX. Proof is very simple as $I_\alpha(P:Q)$ involves the logarithm function. $I_\alpha(P:Q)$ is not a subadditive measure. True to the expectations, it obeys the subadditivity rule for a value of α for a set of probability distributions and for another value, it disobeys the rule for the same set of probability distributions. We have noted our findings in the following

Example 2.3.3.2 : Let $P * Q = \begin{pmatrix} .02 & .08 \\ .27 & .63 \end{pmatrix}$ and $R * S = \begin{pmatrix} .03 & .06 \\ .30 & .61 \end{pmatrix}$ with $P = (.1, .9)$, $Q = (.29, .71)$, $R = (.9, .91)$ and $S = (.33, .67)$.

Then we have $I_\alpha(P * Q:R * S) - I_\alpha(P:R) - I_\alpha(Q:S) = -0.0118943$ for $\alpha = \frac{1}{2}$

and $ = +0.01537$ for $\alpha = 2$

Finally we discuss the convexity property of $I_\alpha(P:Q)$ w.r. to p_i 's and q_i 's in the following

Proposition 2.3.3.5 : $I_\alpha(p_1, \dots, p_n; q_1, \dots, q_n)$ is a convex function of q_i 's if $\alpha > 2$.

Proof : $I_\alpha(P;Q) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^n p_i q_i^{\alpha-1} \right)$

$$\frac{\partial^2 I_\alpha}{\partial q_i^2} = \frac{1}{(1-\alpha)} (\alpha-1) \left(\frac{p_i}{\sum p_i q_i^{\alpha-1}} \right)^2 (\alpha-2) q_i^{\alpha-3} > 0 \text{ for } \alpha > 2.$$

Note that when we are checking $I_\alpha(P;Q)$ for its convexity as a function of q_i 's, we have to treat p_i 's as constants. That proves our claim.

Renyi's measure of inaccuracy is a continuous function of both true probabilities and the asserted probabilities. When $p_i = q_i = \frac{1}{n} \forall i = 1, \dots, n$, $I_\alpha(P;Q)$ is a monotonically increasing function of n . $I_\alpha(P;Q)$ is unchanged if two alternatives about which the same assertion is made are combined. These are three of the four axioms on which Kerridge built his measure and proved the uniqueness of it upto a multiplicative constant. $I_\alpha(P;Q)$ satisfying a majority of these axioms and several other properties that are satisfied by Kerridge's measure becomes almost a perfect measure of inaccuracy.

2.3.4 Rathie and Kannappan Measure of Inaccuracy :

Rathie and Kannappan [47] defined a measure of inaccuracy which corresponds to Havrda-Charvats' measure of information as follows

$$I^\beta(P;Q) = \frac{1}{(2^{\beta-1}-1)} \left\{ \sum_{i=1}^n p_i q_i^{1-\beta} - 1 \right\}. \quad (2.43)$$

It approaches $\frac{-1}{\ln 2} \sum_{i=1}^n p_i \ln q_i$ as $\beta \rightarrow 1$ which is a constant multiple of Kerridge's measure.

$I^\beta(P;Q)$ is a continuous function of both p_i 's and q_i 's, which is fairly obvious. Now we shall verify property II in the following proposition.

Proposition 2.3.4.1 : $I^\beta(P:Q)$ is a monotonically increasing function of n for $p_i = q_i = \frac{1}{n} \forall i = 1, \dots, n$.

Proof :
$$I^\beta(U:U) = \frac{1}{2^{\beta-1}-1} \left\{ \sum_{i=1}^n \frac{1}{n} \left(\frac{1}{n}\right)^{1-\beta} - 1 \right\}$$
$$= \frac{1}{2^{\beta-1}-1} \{n^{\beta-1}-1\} = f(n) \text{ say, then we have}$$

$$f'(n) = \frac{1}{2^{\beta-1}-1} (\beta-1)n^{\beta-2},$$

Case (i) : $\beta > 1$ then $(2^{\beta-1}-1) > 0$ and hence $f'(n) > 0$.

Case (ii) : $\beta < 1$ then $(2^{\beta-1}-1) < 0$ and again $f'(n) > 0$.

$I^\beta(P:Q)$ does not satisfy the third property. But the IV property is satisfied. It is verified in the following

Proposition 2.3.4.2 : $I^\beta(p_1, p_2, p_3; q_1, q_2, q_3) = I^\beta(p_1, p_2+p_3; q_1, q_2)$.

Proof :
$$I^\beta(p_1, p_2, p_3; q_1, q_2, q_3)$$
$$= \frac{1}{(2^{\beta-1}-1)} \{p_1 q_1^{1-\beta} + p_2 q_2^{1-\beta} + p_3 q_3^{1-\beta}\}$$
$$= \frac{1}{(2^{\beta-1}-1)} \{p_1 q_1^{1-\beta} + (p_2+p_3) q_3^{1-\beta}\}$$
$$= I^\beta(p_1, p_2+p_3; q_1, q_2)$$

Remark : Renyi's measure of Inaccuracy can not be defined for P and Q both degenerate distributions, but with non-matching non-zero components. But Rathie-Kannappan's measure can be defined for these distributions also. However in that case

$I^\beta(P:Q) \equiv I^\beta(U_1:U_1) \neq 0$, but $\frac{1}{(1-2^{1-\beta})}$ with this in mind we now prove Vth property in the following

Proposition 2.3.4.3 : $I^\beta(p_1, \dots, p_n; q_1, \dots, q_n) = 0$ if and only if $p_i = q_i = 1$ for some $i = j$ and $p_i = q_i = 0$ for all $i \neq j$.

Proof : The sufficiency part of the proof is obvious. The necessity part of it runs exactly like the proof of Proposition 2.3.3.3, for

$$\frac{1}{2^{\beta-1}-1} \left\{ \sum_{i=1}^n p_i q_i^{1-\beta} - 1 \right\} = 0 \Rightarrow \sum_{i=1}^n p_i (q_i^{1-\beta} - 1) = 0 \text{ which is same as } (2.3.9).$$

Property VI is not satisfied. We have already discussed a situation similar to this property in Remark after the Prop. 2.3.4.2. Property VII is also not satisfied because $I^\beta(P:Q)$ cannot be expressed as a sum of information and directed divergence. With this we now come to deriving the minimum value of $I^\beta(P:Q)$ for a fixed P .

Proposition 2.3.4.4 : The minimum value of $I^\beta(P:Q)$ for a fixed P is attained when

$$q_i = \frac{p_i^{1/\beta}}{\sum_{i=1}^n p_i^{1/\beta}} \text{ and it is given by } \frac{1}{2^{\beta-1}-1} \left\{ \left(\sum_{i=1}^n p_i^{1/\beta} \right)^\beta - 1 \right\}. \text{ Moreover as } \beta \text{ approaches } 1, \text{ Min}_Q I^\beta(P:Q) \text{ approaches } \frac{-\sum_{i=1}^n p_i \ln p_i}{\ln 2}.$$

Proof : $I^\beta(P:Q)$ is convex with respect to q_i 's for $\beta > 1$.

Making use of this fact and applying Lagrange's multipliers method we minimize $I^\beta(P:Q)$ subject to $\sum_{i=1}^n q_i = 1$, to get

$$q_i = \frac{p_i^{1/\beta}}{\sum_{i=1}^n p_i^{1/\beta}} \quad (2.44)$$

as the minimizing p.distribn. The minimum value of $I^\beta(P:Q)$ is obtained by substituting (2.44) in (2.43) as

$$\begin{aligned} \min_Q I^\beta(P:Q) &= I^\beta(P:Q) = \frac{1}{2^{\beta-1}-1} \left\{ \sum_{i=1}^n p_i \left(\frac{p_i^{1/\beta}}{\sum_{i=1}^n p_i^{1/\beta}} \right)^{1-\beta} - 1 \right\} \\ &= \frac{1}{2^{\beta-1}-1} \left\{ \left(\sum_{i=1}^n p_i^{1/\beta} \right)^\beta - 1 \right\}. \end{aligned} \quad (2.45)$$

Now we shall show that $I^\beta(P:Q) \rightarrow -(\sum_{i=1}^n p_i \ln p_i) / \ln 2$.

$$\begin{aligned} \lim_{\beta \rightarrow 1} I^\beta(P:Q) &= \lim_{\beta \rightarrow 1} \frac{1}{2^{\beta-1}-1} \left\{ \left(\sum_{i=1}^n p_i^{1/\beta} \right)^\beta - 1 \right\} \\ &= \lim_{\beta \rightarrow 1} \left\{ \frac{-(\sum_{i=1}^n p_i^{1/\beta})^{\beta-1}}{\beta} - \sum_{i=1}^n p_i^{1/\beta} \ln p_i \right. \\ &\quad \left. + \left(\sum_{i=1}^n p_i^{1/\beta} \right)^\beta \ln \left(\sum_{i=1}^n p_i^{1/\beta} \right) \right\} \\ &= -(\sum_{i=1}^n p_i \ln p_i) / \ln(2). \end{aligned}$$

Let P, Q, R and S be any probability distributions such that P and Q and R and S are mutually independent.

Then we have

$$\begin{aligned} I^\beta(PQ:RS) &= \frac{1}{(2^{\beta-1}-1)} \left\{ \sum_{j=1}^m \sum_{i=1}^n (p_i q_j) (r_i s_j)^{\beta-1} - 1 \right\} \\ &= \frac{1}{2^{\beta-1}-1} \left\{ \sum_{i=1}^n p_i r_i^{\beta-1} \sum_{j=1}^m q_j s_j^{\beta-1} - 1 \right\} \\ &= \frac{1}{2^{\beta-1}-1} \left\{ \sum_{i=1}^n p_i r_i^{\beta-1} - 1 \right\} + \frac{1}{2^{\beta-1}-1} \left\{ \sum_{j=1}^m q_j s_j^{\beta-1} - 1 \right\} \\ &\quad - \frac{1}{2^{\beta-1}-1} \left\{ \sum_{i=1}^n p_i r_i^{\beta-1} - 1 \right\} \left\{ \sum_{j=1}^m q_j s_j^{\beta-1} - 1 \right\} \end{aligned}$$

$$= I^\beta(P:R) + I^\beta(Q:S) - (2^{\beta-1} - 1)^{-1} I^\beta(P:R) I^\beta(Q:S) \quad (2.46)$$

From (2.46) we can infer that $I^\beta(P:Q)$ is not additive. Therefore $I^\beta(P:Q)$ does not satisfy property IX for all values of β .

$I^\beta(P:Q)$ is neither subadditive nor is superadditive monotonously. We deduce that in the following example.

Example 2.3.4.1 : Consider the same distributions as we did in Example 2.3.3.2. Then we get

$$\begin{aligned} I^{1/2}(P * Q:R * S) - I^{1/2}(P:R) - I^{1/2}(Q:S) &= 1.1506211 - 1.2417572 \\ &= -0.0911361 \end{aligned}$$

and

$$\begin{aligned} I^2(P * Q:R * S) - I^2(P:R) - I^2(Q:S) &= 2.9327869 - 2.0386115 \\ &= 0.8941754 \end{aligned}$$

So the conclusion is that $I^\beta(P:Q)$ is subadditive for $\alpha = \frac{1}{2}$ and superadditive for $\alpha = 2$ for the distributions given in example (2.39).

We shall now consider convexity of $I^\beta(P:Q)$ w.r.to q_i 's. We have from (2.38)

$$\begin{aligned} \frac{\partial I^\beta(P:Q)}{\partial q_i} &= \frac{\partial}{\partial q_i} \frac{1}{2^{-1+\beta}-1} \left\{ \sum_{i=1}^n p_i q_i^{1-\beta} - 1 \right\} \\ &= \frac{1}{2^{\beta-1}-1} p_i (-\beta+1) q_i^{-\beta} \quad \text{and} \end{aligned}$$

$$\frac{\partial^2 I^\beta}{\partial q_i^2} = \frac{1}{2^{\beta-1}-1} p_i (-1+\beta)\beta q_i^{-\beta-1} > 0 \quad \text{for all } \beta > 0.$$

Therefore we conclude that $I^\beta(P:Q)$ is a convex function of q_i^S .

We conclude this subsection with a brief review of the properties possessed by Rathie-Kannappan measure of inaccuracy, $I^\beta(P:Q)$. It is a continuous function of both true and asserted probabilities. $I^\beta(U:U)$ is an increasing function of n , the number of components of U . It does not satisfy the recursivity property but it is a convex function of q_i 's. It is neither additive nor is subadditive. It attains its minimum value for

$\{q_i\}_1^n = \left\{ \frac{p_i^{1/\beta}}{\sum p_i^{1/\beta}} \right\}_1^n$ and this minimum value tends to

$-\left(\sum_{i=1}^n p_i \ln p_i / \ln 2 \right)$ as $\beta \rightarrow 1$, expectedly. This minimum value is zero iff $p_i = q_i = 1$ for some i and $p_i = q_i = 0$ for all other values of i . This measure deviates from Renyi's measure in only one property, the additivity property.

2.3.5 Sharma and Guptas' Measures of Inaccuracy :

Sharma and Gupta have proposed [59] the following measures of inaccuracy of two or more parameters.

(a) Log-Measure :

$$I_{\alpha, \beta}^L(P:Q) = -2^\beta \sum_{i=1}^n p_i^\alpha q_i^\beta \log q_i \quad \alpha > 0, \beta \geq 0 \quad (2.47)$$

(b) Power-Measure :

$$I_{\alpha, \beta}^P(P:Q) = (2^{-\beta} - 2^{-\gamma})^{-1} \sum_{i=1}^n p_i^\alpha (q_i^\beta - q_i^\gamma) \quad \alpha > 0, \beta, \gamma \geq 0, \beta \neq \gamma \quad (2.48)$$

(c) Sine-Measure :

$$I_{\alpha, \beta}^S(P:Q) = \frac{-2^\beta}{\sin} \sum_{i=1}^n p_i^\alpha q_i^\beta \sin(\gamma \log q_i) \quad \alpha > 0, \beta \geq 0, \gamma \neq 0 \quad (2.49)$$

(a) Log-Measure

From the definition we can easily note that $I_{\alpha,\beta}^L(P:Q)$ is a continuous function of p_i 's and q_i 's. Therefore it satisfies the property I.

Now we shall verify the second property in the following

Proposition 2.3.5.1 : $I_{\alpha,\beta}^L(P:Q)$ is a monotonically increasing function of n for $p_i = q_i = \frac{1}{n} \forall i = 1, \dots, n$ iff

$n > \exp\{1/(-\alpha-\beta+1)\}$ for $\alpha+\beta < 1$ etc.

Proof : We have

$$\begin{aligned} I_{\alpha,\beta}^L(U:U) &= -2^\beta n \left(\frac{1}{n}\right)^\alpha \left(\frac{1}{n}\right)^\beta \ln \left(\frac{1}{n}\right) \\ &= 2^\beta n^{1-\alpha-\beta} \ln n = f(n), \text{ say, then} \end{aligned}$$

$$f'(n) = 2^\beta n^{-(\alpha+\beta)} \{(1-\alpha-\beta) \ln n - 1\} \quad (2.50)$$

For $f'(n) > 0$ we must have $(1-\alpha-\beta) \ln n - 1 > 0$ because $n^{-\alpha-\beta}$ is always > 0 and so is 2^β .

Case (i) : Let $\alpha+\beta < 1$ then we have $f'(n) > 0$ iff

$$(1-\alpha-\beta) \ln n - 1 > 0$$

$$\text{or } \ln n > \frac{1}{1-\alpha-\beta}$$

$$\text{or } n > \exp\{1/(1-\alpha-\beta)\}.$$

Case (ii) : Let $\alpha+\beta > 1$ then we have $f'(n) > 0$ if and only if

$$(1-\alpha-\beta) \ln n - 1 > 0$$

$$\text{or } (1-\alpha-\beta) \ln n > 1$$

$$\text{or } \ln n < \frac{1}{(1-\alpha-\beta)}$$

$$\text{or } n < \exp\{1/(1-\alpha-\beta)\}.$$

We shall illustrate the situation in the following

Example 2.3.5.1 : Let $\alpha = 0.5$ and $\beta = 1.5$ then we take up the Case (ii) of Proposition 2.3.5.1 i.e., we have

$$n < \exp\{1/(1-\alpha-\beta)\} = \exp\{-1\} = e^{-1} < 1$$

which is impossible because we shall want n to be only a positive integer greater than or equal to 2.

Now consider $\alpha = 0.5$ and $\beta = 0.4$ then we pertain to Case (i) of Proposition 2.3.5.1 i.e., we have

$$n > \exp\{1/(1-\alpha-\beta)\} = e^{(1/0.1)} = e^{10}.$$

Now consider $n = 10$ lets say. Then we have from (2.50)

$$f'(n) = 2^{0.4} (10^{-0.9}) \{ (1-0.9) \ln 10 - 1 \} = 2^{0.4} 10^{-0.9} \{-0.76\}$$

$$= -0.12624 < 0$$

Through this example we realize the fact that in any practical situation this measure may not always be a monotonically increasing function of n for $p_i = q_i = \frac{1}{n} \forall i = 1, \dots, n$.

Property IV states that the inaccuracy of a statement is unchanged if two alternatives about which the assertion is made are combined. Let us consider

$$\begin{aligned} I_{\alpha, \beta}^L(P:Q) &\equiv I_{\alpha, \beta}^L(p_1, p_2, p_3, q_1, q_2, q_2) \\ &= -2^\beta \{ p_1^\alpha q_1^\beta \ln q_1 + p_2^\alpha q_2^\beta \ln q_2 + p_3^\alpha q_2^\beta \ln q_2 \} \end{aligned}$$

$$= -2^{\beta} \{ p_1^{\alpha} q_1^{\beta} \ln q_1 + (p_2^{\alpha} + p_3^{\alpha}) q_2^{\beta} \ln q_2 \}$$

$$\neq I_{\alpha, \beta}^L(p_1, p_2 + p_3; q_1, q_2) \quad \text{unless } \alpha = 1.$$

So property IV is not satisfied for $I_{\alpha, \beta}^L(P:Q)$ unless $\alpha = 1$.

Note that for $\alpha = 1$ and $\beta = 0$ this measure happens to be Kerridge measure of inaccuracy and Kerridge's measure satisfies property IV.

We shall now consider property V which states that

$I(P:Q)$ should be zero iff $p_i = q_i = 1$ for some i and $p_i = q_i = 0$ for all other i . Obviously $I_{\alpha, \beta}^L(D_i:D_i) = 0$ with the convention that $0 \ln 0 = 0$ where D_i is the degenerate distribution with i^{th} component being unity. We can also see that $I_{\alpha, \beta}^L(P:Q)$ approaches infinity if for some i , $q_i = 0$ and $p_i \neq 0$. Now we shall discuss the minimum of this measure. Because $I_{\alpha, \beta}^L(P:Q)$ is always positive and it is zero iff $P = Q = D_i$ for some $i = 1, \dots, n$, we derive that the minimum of this measure is zero. This measure is not additive but $I_{\alpha, \beta}^L(P_1 P_2:Q_1 Q_2)$ can be expressed as a weighted sum of $I_{\alpha, \beta}^L(P_1:Q_1)$ and $I_{\alpha, \beta}^L(P_2:Q_2)$. For,

$$\begin{aligned} I_{\alpha, \beta}^L(P_1 P_2:Q_1 Q_2) &= -2^{\beta} \sum_{i=1}^n \sum_{j=1}^m ((p_i^1 p_j^2)^{\alpha} (q_i^1 q_j^2)^{\beta} \ln (q_i^1 q_j^2)) \\ &= -2^{\beta} \sum_{i=1}^n p_i^1 q_i^1{}^{\beta} \ln q_i^1 (\sum_{j=1}^m p_j^2 q_j^2{}^{\alpha})^{\beta} \\ &\quad - 2^{\beta} \sum_{j=1}^m (p_j^2)^{\alpha} (q_j^2)^{\beta} \ln q_j^2 (\sum_{i=1}^n p_i^1 q_i^1{}^{\alpha})^{\beta} \\ &= (\sum_{j=1}^m p_j^2 q_j^2{}^{\alpha} 2^{\beta}) I_{\alpha, \beta}^L(P_1:Q_1) + (\sum_{i=1}^n p_i^1 q_i^1{}^{\alpha} 2^{\beta}) I_{\alpha, \beta}^L(P_2:Q_2). \end{aligned}$$

This measure is subadditive according to its authors. It is not convex w.r. to q_i 's, See Kapur [24].

On considering this measure with an over all view, we conclude that its positive aspects are very limited. It is always positive and is zero only for degenerate distributions with matching non-zero components. Its desired highpoint might have been the subadditivity, but the property itself is not highly desirable for an inaccuracy measure. $I_{\alpha,\beta}^L(P:Q)$ does not satisfy any of the important properties like convexity, monotonically increasing of $\frac{d}{dn} I_{\alpha,\beta}^L(U:U)$, etc.

(b) Sharma and Guptas' Power and Sine Measure of Inaccuracy :

One can easily see that $I_{\alpha,\beta}^P(P:Q)$ and $I_{\alpha,\beta,\gamma}^S(P:Q)$ are continuous functions of p_i 's and q_i 's. But only $I_{\alpha,\beta}^P(P:Q)$ is always positive. It can easily be verified $I_{\alpha,\beta,\gamma}^S(P:Q)$ can take negative values also. Consider the following

Example 2.3.5.2 : Let $\alpha = 1$, $\beta = 1$, $\gamma = -\frac{\pi}{2}$ and

$P = (0.1, 0.9)$ and $Q = (0.6, 0.4)$. Then we have

$$I_{\alpha,\beta,\gamma}^S(P:Q) = -0.80006 < 0.$$

The amount of inaccuracy must be always positive or at the least, zero. But it is difficult to give any kind of interpretation to inaccuracy being negative. Sharma and Guptas' power measure violates this most basic property.

We shall now consider $I_{\alpha,\beta,\gamma}^P(U:U)$ and $I_{\alpha,\beta,\gamma}^S(U:U)$ as

functions of n , $f(n)$ and $g(n)$ and derive conditions for $f'(n) > 0$ and $g'(n) > 0$.

Proposition 2.3.5.2 : $f(n) \equiv I_{\alpha, \beta, \gamma}^P(U:U)$ is a monotonically increasing function of n if and only if

(i) $\beta < \gamma$

(a) $(\alpha + \beta) < 1$, $n > \left\{ \frac{1-\alpha-\gamma}{1-\alpha-\beta} \right\}^{\frac{1}{\gamma-\beta}}$, (b) $(\alpha + \beta) > 1$, $n > \left\{ \frac{1-\alpha-\beta}{1-\alpha-\gamma} \right\}^{\frac{1}{\beta-\gamma}}$

(ii) $\beta \geq \gamma$

(c) $(\alpha + \beta) < 1$, $n > \left\{ \frac{1-\alpha-\beta}{1-\alpha-\gamma} \right\}^{\frac{1}{\beta-\gamma}}$, (d) $(\alpha + \beta) > 1$, $n > \left\{ \frac{1-\alpha-\gamma}{1-\alpha-\beta} \right\}^{\frac{1}{\gamma-\beta}}$

Proof : From (4.3.2) we have

$$f(n) \equiv I_{\alpha, \beta, \gamma}^P(U:U) = \frac{1}{2^{-\beta} - 2^{-\gamma}} \{n^{1-\alpha-\beta} - n^{1-\alpha-\gamma}\}$$

Then we have

$$f'(n) = \frac{n^{-\alpha}}{(2^{-\beta} - 2^{-\gamma})} \{(1-\alpha-\beta)n^{-\beta} - (1-\alpha-\gamma)n^{-\gamma}\} \quad (2.51)$$

In RHS of the above expression, $n^{-\alpha}$ is always positive. Therefore we consider the following two cases :

Case (1) : $\beta < \gamma \implies 2^{-\beta} > 2^{-\gamma}$ or $(2^{-\beta} - 2^{-\gamma}) > 0$. Therefore for $f'(n)$ to be positive, we must have

$$(1-\alpha-\beta)n^{-\beta} - (1-\alpha-\gamma)n^{-\gamma} > 0$$

$$\text{or } (1-\alpha-\beta)n^{-\beta} > (1-\alpha-\gamma)n^{-\gamma}$$

$$\text{or } (1-\alpha-\beta)n^{\gamma-\beta} > (1-\alpha-\gamma).$$

(2.52)

There are further two cases.

Case (a) : $1 > \alpha + \beta \implies (1 - \alpha - \beta) > 0$. Then (2.52) becomes

$$n^{\gamma - \beta} > \left\{ \frac{1 - \alpha - \gamma}{1 - \alpha - \beta} \right\} \text{ for } f'(n) > 0$$

$$\text{or } n > \left\{ \frac{1 - \alpha - \gamma}{1 - \alpha - \beta} \right\}^{\frac{1}{\gamma - \beta}} \text{ because } (\gamma - \beta) > 0 \quad (2.53)$$

Case (b) : $1 < \alpha + \beta \implies (1 - \alpha - \beta) < 0$. Then (2.52) becomes

$$n^{\gamma - \beta} < \left\{ \frac{1 - \alpha - \gamma}{1 - \alpha - \beta} \right\} \text{ for } f'(n) > 0$$

$$\text{or } n^{\beta - \gamma} > \left\{ \frac{1 - \alpha - \beta}{1 - \alpha - \gamma} \right\}$$

$$\text{or } n < \left\{ \frac{1 - \alpha - \beta}{1 - \alpha - \gamma} \right\}^{\frac{1}{\beta - \gamma}} \text{ because } \beta < \gamma \quad (2.54)$$

Case (ii) : $\beta \geq \gamma \implies 2^{-\beta} \leq 2^{-\gamma}$ or $(2^{-\beta} - 2^{-\gamma}) \leq 0$. Therefore for $f'(n)$ to be positive, from (2.51) we must have

$$(1 - \alpha - \beta)n^{-\beta} - (1 - \alpha - \gamma)n^{-\gamma} < 0$$

$$\text{or } (1 - \alpha - \beta)n^{\gamma - \beta} < (1 - \alpha - \gamma).$$

There are again two cases further.

Case (c) : $(1 - \alpha - \beta) > 0$. In this case for $f'(n) > 0$ we must have

$$n^{\gamma - \beta} < \left\{ \frac{1 - \alpha - \gamma}{1 - \alpha - \beta} \right\}$$

$$\text{or } n^{\beta - \gamma} > \left\{ \frac{1 - \alpha - \beta}{1 - \alpha - \gamma} \right\}$$

$$\text{or } n > \left\{ \frac{1 - \alpha - \beta}{1 - \alpha - \gamma} \right\}^{\frac{1}{\beta - \gamma}} \quad (\beta - \gamma) \geq 0 \quad (2.55)$$

Case (d) : $(1-\alpha-\beta) < 0$. In this case for $f'(n) > 0$ we must have

$$n^{\gamma-\beta} > \left\{ \frac{1-\alpha-\gamma}{1-\alpha-\beta} \right\}$$

$$\text{or } n < \left\{ \frac{1-\alpha-\gamma}{1-\alpha-\beta} \right\}^{\frac{1}{\gamma-\beta}} \quad \text{because } (-\beta+\gamma) \leq 0 \quad (2.56)$$

Therefore we have conditions for $f'(n)$ to be positive in (2.53), (2.54), (2.55) and (2.56).

We can illustrate our conditions in the following example.

Example 2.3.5.3 : Let $\alpha = 0.5$, $\beta = 1$, $\gamma = 2$. Then $\beta \leq \gamma$ and $(1-\alpha-\beta) < 0$. Therefore case (b) appeals here. We must have for $f'(n) > 0$

$$n < \left\{ \frac{1-\alpha-\beta}{1-\alpha-\gamma} \right\}^{\frac{1}{\beta-\gamma}} = \left\{ \frac{-0.5}{-1.5} \right\}^{-1} = 3.$$

Now let us consider $n = 2$. Then we have

$$f'(2) = 2.828428 \{-0.5 \ 0.5 + 1.5 \ 0.25\} = 0.353535 > 0.$$

Now let us consider $n = 4$. Then we have

$$f'(4) = 2 \{-0.5 \ 0.25 + 1.5 \ 0.0625\} = -0.0625 < 0.$$

Therefore for the above set of values of α, β, γ, n can be equal to atmost 2.

Example 2.3.5.4 : Let $\alpha = 0.4$, $\beta = 0.5$, $\gamma = 0.3$. This is case (c) of Proposition 2.3.5.2. For $f'(n) > 0$ we must have

$$n > \left\{ \frac{1-\alpha-\beta}{1-\alpha-\gamma} \right\}^{\frac{1}{\beta-\gamma}} = \left\{ \frac{0.1}{0.3} \right\}^{\frac{1}{0.2}} = 0.0041. \quad \text{That is in this case } n \text{ can be any positive integer.}$$

Let $n = 4$ then we have

$$f'(4) = -5.46 \times \{0.1 \times 0.5 - 0.3 \times 0.659754\} = 0.80767 > 0.$$

Proposition 2.3.5.3 : $g(n) \equiv I_{\alpha, \beta, \gamma}^S(U:U)$ is a monotonically increasing function of n if and only if $\sin \gamma$ and

$\{\gamma \cos(\gamma \ln n) + (1-\alpha-\beta) \sin(\gamma \ln n)\}$ are both of the same sign.

Proof : $g(n) = \frac{2^\beta}{\sin \gamma} \{n^{1-\alpha-\beta} \sin(\gamma \ln n)\}$ and on differentiating w.r. to n , we get

$$g'(n) = \frac{n^{-\alpha-\beta} 2^\beta}{\sin \gamma} \{\gamma \cos(\gamma \ln n) + (1-\alpha-\beta) \sin(\gamma \ln n)\} \quad (2.57)$$

From (2.57) it is evident that the condition is necessary and sufficient for $g'(n)$ to be greater than zero.

Example 2.3.5.5 : Let $\gamma = \frac{\pi}{2}$, $\alpha = 0.5$, $\beta = 0$.

Then we have $\sin \gamma = \sin \frac{\pi}{2} = 1$ and now with $n = 3$ we have

$$\{\gamma \cos(\gamma \ln n) + (1-\alpha-\beta) \sin(\gamma \ln n)\} = \{-\frac{\pi}{2} \times 0.15428 + 0.5 \times 0.49401\} \\ = 0.0046625 > 0$$

there by giving us $g'(n) > 0$.

Now with the same set of values of α, β, γ we take $n = 6$ to find that $\{\gamma \cos(\gamma \ln n) + (1-\alpha-\beta) \sin(\gamma \ln n)\} = -1.3270441 < 0$.

Here $\sin \gamma$ and $\{\gamma \cos(\gamma \ln n) + (1-\alpha-\beta) \sin(\gamma \ln n)\}$ have different signs. Hence $g'(n) < 0$ and $g(n)$ is not increasing at $n = 6$.

We can easily see from the periodic nature of $I_{\alpha,\beta,\gamma}^S(P:Q)$ that for any set of values of α, β, γ , we will have several values of n for which $g'(n) > 0$ and several others for which $g'(n) < 0$. Therefore it is not possible to obtain any condition on n so that $g'(n) > 0$ is satisfied.

We shall now discuss other properties of these two measures of inaccuracy.

Neither $I_{\alpha,\beta,\gamma}^P(P:Q)$ nor $I_{\alpha,\beta,\gamma}^S(P:Q)$ satisfy the recursivity property.

$I_{\alpha,\beta,\gamma}^P(P:Q)$ and $I_{\alpha,\beta,\gamma}^S(P:Q)$ also violate property IV. That is the inaccuracy (measured by $I_{\alpha,\beta,\gamma}^P(P:Q)$ or $I_{\alpha,\beta,\gamma}^S(P:Q)$) of a statement changes when two outcomes regarding which same assertion is made are combined. This highly desirable property is not satisfied by any of Sharma and Guptas' three measures. They are not additive and are not convex. We can not find the minimum value of any of these measures for fixed $\{p_i\}$ as they are not convex and hence Lagrange's method is inconclusive. But for $I_{\alpha,\beta,\gamma}^P(P:Q)$ the minimum is zero which is attained by $P = Q = D_i$ where D_i is the degenerate distribution with i^{th} component being unity. It is zero only for the degenerate distributions. $I_{\alpha,\beta,\gamma}^S(P:Q)$ takes negative values also as is seen in Example 2.3.5.2 and so its minimum value can not be obtained.

2.3.6 Kapur's Measures of Inaccuracy :

Kapur proposed a very generalized measure of inaccuracy [23]

which contains many known measures of inaccuracy as special cases. It is defined as follows :

$$I_K(P:Q) = \{(\alpha-1)k\}^{-1} \{n^{-b} (\sum_{i=1}^n p_i^\alpha)^a (\sum_{i=1}^n q_i^{1-\alpha})^b (\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha})^c \\ (\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha})^k - 1\} + n^{b+c+(\alpha-1)k} \{1 - n^{-k(\alpha-1)} \\ - (\sum_{i=1}^n p_i^\alpha)^{a+c+k} + (\sum_{i=1}^n p_i^\alpha)^{a+c} n^{-k(\alpha-1)}\} \quad (2.58)$$

This measure is built to satisfy the following properties :

- (i) $I(P:Q) \geq 0$
- (ii) $I(P:Q) \geq I(P:P) = H(P)$
- (iii) $I(P:Q) = H(P)$ if and only if $Q = P$.

But the special cases of this measure are of importance. We now present a list of special cases of (2.58).

- (i) If $a = b = c = 0$ and $k = 1$ and $\alpha \rightarrow 1$ we get Kerridge's measure of inaccuracy, $-\sum_{i=1}^n p_i \ln q_i$.

- (ii) If $a = b = c = 0$, $k = 1$, we get another measure of inaccuracy due to Kapur [23] which is given by

$$I_{K_2}^\alpha(P:Q) = \frac{1}{\alpha-1} \left(\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} - 1 \right) + \frac{n^{\alpha-1}}{1-\alpha} \left(\sum_{i=1}^n p_i^\alpha - 1 \right) \quad (2.59)$$

- (iii) If $a = b = c = 0$, $k \rightarrow 0$ we get the measure which is obtained individually by Kapur and J.A. Van der Lubbe [60],

$$I_{L1}^\alpha(P:Q) = \frac{1}{\alpha-1} \ln \left\{ \frac{\sum_{i=1}^n p_i^\alpha q_i^{1-\alpha}}{\sum_{i=1}^n p_i^\alpha} \right\} \quad (2.60)$$

(iv) If $a = b = c = 0$, $k = -1$ we get the following new measure of inaccuracy

$$I_{KN}^{\alpha}(P:Q) = \frac{1}{\alpha-1} \left\{ 1 - \frac{1}{n \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}} \right\} + \frac{n^{1-\alpha}}{1-\alpha} \left\{ 1 - \frac{1}{n \sum_{i=1}^n p_i^{\alpha}} \right\} \quad (2.61)$$

(v) If $a = 1$, $b = 0$, $c = 0$, $k = -1$, we get Lubbe's [60] measure of inaccuracy

$$I_{L2}^{\alpha}(P:Q) = \frac{1}{\alpha-1} \left\{ 1 - \left(\sum_{i=1}^n p_i^{\alpha} / \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} \right) \right\}. \quad (2.62)$$

Now we shall first undertake to study the properties of the measures defined in (2.59) - (2.62). Later we shall consider the general measure (2.58).

2.3.6.1 Kapur's Measure of Inaccuracy :

$$I_{K2}^{\alpha}(P:Q) = \frac{1}{\alpha-1} \left(\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} - 1 \right) + \frac{n^{\alpha-1}}{1-\alpha} \left(\sum_{i=1}^n p_i^{\alpha} - 1 \right).$$

It is obviously a continuous function of both p_i 's and q_i 's. Now let us consider

$$\begin{aligned} I_{K2}^{\alpha}(U:U) &= \frac{1}{\alpha-1} (n^{1-\alpha} \cdot n^{\alpha-1} - 1) + \frac{n^{\alpha-1}}{1-\alpha} (n^{1-\alpha} - 1) \\ &= \frac{1}{\alpha-1} (0) - \frac{n^{\alpha-1}}{1-\alpha} + \frac{1}{1-\alpha} = \frac{1}{1-\alpha} \{1 - n^{\alpha-1}\} = f(n) \text{ (say)} \end{aligned} \quad (2.63)$$

We then have $f'(n) = n^{\alpha-2}$ which is always greater than or equal to zero. Therefore we conclude that

Proposition 2.3.6.1 : When n equally likely outcomes are stated to be equally likely then $I_{K2}^{\alpha}(P:Q)$ is a monotonically increasing function of n .

We could not find any recursive relation for $I_{K2}^{\alpha}(P:Q)$. Actually one can observe that the first part of this measure (2.59) is Havrda-Charvats' directed divergence $D^{\alpha}(P:Q)$ and the second part is $(n^{\alpha-1})$ times Havrda-Charvats' information measure $H^{\alpha}(P)$. As we had seen in previous sections, no measure which has an exponent of the true probabilities in its expression viz. $(\dots p_i^{\alpha} \dots)$ satisfies property IV which combines two outcomes when they are asserted to have same probabilities. Hence $I_{K2}^{\alpha}(P:Q)$ also does not satisfy Property IV.

$I_{K2}^{\alpha}(P:Q) = 0$ iff $P = Q$ for some i . It is very easy to see this fact. $I_{K2}^{\alpha}(P:Q)$ does not approach infinity if a $q_i = 0$ and the corresponding $p_i \neq 0$. Now we shall consider the minimum value of $I_{K2}^{\alpha}(P:Q)$ for fixed $\{p_i\}_{i=1}^n$. Since it can be expressed as a sum of an information measure and a directed divergence measure, $I_{K2}^{\alpha}(P:Q)$ satisfies both properties VI and VII. Thus for a fixed $\{p_i\}$, $p_i = q_i$ gives a minimum of $I_{K2}^{\alpha}(P:Q)$ which given by

$$\min_Q \{I_{K2}^{\alpha}(P:Q)\} = \frac{n^{\alpha-1}}{1-\alpha} \left(\sum_{i=1}^n p_i^{\alpha} - 1 \right) \text{ which tends to } - \sum_{i=1}^n p_i \ln p_i$$

as $\alpha \rightarrow 1$.

If the variations of both P and Q are considered $p_i = q_i$, $\forall i = 1, \dots, n$ is a minimax point for $I_{K2}^{\alpha}(P:Q)$.

Because Havrda-Charvats' both measures (i.e., of information and directed divergence) are non-additive, we can easily deduce that $I_{K2}^{\alpha}(P:Q)$ is non-additive. Similarly $I_{K2}^{\alpha}(P:Q)$ is not subadditive.

$I_{K2}^{\alpha}(P:Q)$ is convex w.r. to q_i 's for $\alpha < 2$ because sum of two convex functions is convex and Havrda-Charvat's measures are convex for $\alpha < 2$.

2.3.6.2 Lubbe's Measure of Inaccuracy :

It is defined as

$$I_{L1}^{\alpha}(P:Q) = \frac{1}{\alpha-1} \ln \left\{ \frac{\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}}{\sum_{i=1}^n p_i^{\alpha}} \right\}$$

$$= \frac{1}{\alpha-1} \ln \left\{ \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} \right\} + \frac{1}{1-\alpha} \ln \left\{ \sum_{i=1}^n p_i^{\alpha} \right\} \quad (2.64)$$

We observe from the above expression that Lubbe's measure of inaccuracy is a sum of Renyi's measure of directed divergence $D_{\alpha}(P:Q)$ and Renyi's measure of information $H_{\alpha}(P)$.

We can easily see that $I_{L1}^{\alpha}(P:Q)$ is a continuous function of both p_i 's and q_i 's. Now let $P = Q = U = (\frac{1}{n}, \dots, \frac{1}{n})$ in (4.4.7). Then we have

$$I_{L1}^{\alpha}(U:U) = \frac{1}{\alpha-1} \ln \{ n^{1-\alpha} n^{\alpha-1} \} + \frac{1}{1-\alpha} \ln \{ n^{1-\alpha} \}$$

$= \ln n$ which is an increasing function of n . Infact

we could have directly stated that $I_{L1}^{\alpha}(U:U)$ is an increasing fn.

of n from (2.64) as we know that $D_\alpha(P:Q) = 0$ if $P = Q$ and $H_\alpha(U) = \ln n$. Property II is satisfied by $I_{L1}^\alpha(P:Q)$.

Properties III and IV are not satisfied by $I_{L1}^\alpha(P:Q)$ which can be seen easily.

$I_{L1}^\alpha(D_i:D_i) = 0 \forall i = 1, \dots, n$ but $I_{L1}^\alpha(P:Q)$ does not approach infinity if $q_i = 0$ and the corresponding $p_i = 0$ unlike the Kerridge's measure. It has a finite value. Now we shall find the minimum value of $I_{L1}^\alpha(P:Q)$.

Proposition 2.3.6.2 : For a fixed $\{p_i\}$, $I_{L1}^\alpha(P:Q)$ is minimum when $q_i = p_i \forall i = 1, \dots, n$. This value of the inaccuracy is the amount of uncertainty involved in the probability distribution $\{p_i\}_{i=1}^n$.

Proof : Using the fact that $I_{L1}^\alpha(P:Q) = D_\alpha(P:Q) + H_\alpha(P)$ from (2.64) and that $D_\alpha(P:Q)$ is a convex function of q_1, \dots, q_n we apply Lagrange's method to obtain the $I_{L1}^\alpha(P:Q)$ minimizing distribution $Q = \{q_i\}$ as

$$q_i = p_i \quad \forall i = 1, \dots, n$$

$$\text{and } \min_Q \{I_{L1}^\alpha(P:Q)\} = I_{L1}^\alpha(P:Q) = I_{L1}^\alpha(P) = 0 + \frac{1}{1-\alpha} \ln p_i^\alpha = H_\alpha(P).$$

If we consider the variations of both $\{p_i\}$ and $\{q_i\}$, the minimum value of $I_{L1}^\alpha(P:Q)$ is given by $P = Q = D_i$ for any $i = 1, \dots, n$, $p_i = q_i = \frac{1}{n} \forall i = 1, \dots, n$ gives a minimax point for $I_{L1}^\alpha(P:Q)$ as for this, the first term of (2.64) is zero and the second term attains its maximum value, namely $\ln n$.

Thus we have seen that $I_{L1}^{\alpha}(P:Q)$ satisfies Properties VII and VIII. Now we shall show that $I_{L1}^{\alpha}(P:Q)$ is additive. It is because from (2.64)

$$I_{L1}^{\alpha}(P:Q) = D_{\alpha}(P:Q) + H_{\alpha}(P)$$

and both $D_{\alpha}(P:Q)$ and $H_{\alpha}(P)$ are additive functions of their arguments. However $I_{L1}^{\alpha}(P:Q)$ is not subadditive.

$I_{L1}^{\alpha}(P:Q)$ is a convex function of Q . This fact can easily be verified if (2.64) and the convexity of $D_{\alpha}(P:Q)$ is considered, for $0 < \alpha < 2$. For $\alpha > 2$ it is concave.

We conclude this section by observing that $I_{L1}^{\alpha}(P:Q)$ violates only the recursivity property and the property that the inaccuracy of a statement should be infinity if an outcome is asserted to have zero probability whereas in reality it has a positive probability and that $I_{L1}^{\alpha}(P:Q)$ is convex function of q_i 's for only $0 < \alpha < 2$. Otherwise it satisfies all the other properties satisfied by the Kerridge's measure of inaccuracy and hence can be termed as a measure which is as good as Renyi's measure of inaccuracy.

2.3.6.3 Kapur's New Measure of Inaccuracy :

It is defined as follows :

$$I_{KN}^{\alpha}(P:Q) = \frac{1}{\alpha-1} \left\{ 1 - \frac{1}{n \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}} \right\} + \frac{n^{1-\alpha}}{1-\alpha} \left\{ 1 - \frac{1}{n \sum_{i=1}^n p_i^{\alpha}} \right\} \quad (2.65)$$

Remark : We point out that this measure can not accept the set of distributions $P = D_1$ and $Q = D_2$ where D_1 and D_2 are degenerate

distributions with 1st and 2nd components being unity respectively, because the 1st term is not defined for them. Except for the degenerate distributions with non-matching non-zero components, $I_{KN}^{\alpha}(P:Q)$ is a continuous function of both p_i 's and q_i 's. We have same drawback to Kerridge's and Rathie-Kannappan's measures of inaccuracy.

Proposition 2.3.6.3 : $I_{KN}^{\alpha}(P:Q)$ is an increasing function of n if n equally likely outcomes are stated equally likely.

Proof: We have $I_{KN}^{\alpha}(U:U) = \frac{1}{\alpha-1} \left\{ 1 - \frac{1}{n^{1-\alpha}} \frac{1}{n^{\alpha-1}} \right\} + \frac{n^{1-\alpha}}{1-\alpha} \left\{ 1 - \frac{1}{n^{1-\alpha}} \right\}$

$$= \frac{n^{1-\alpha}-1}{(1-\alpha)} = f(n) \text{ (say) then we have}$$

$$f'(n) = n^{-\alpha} > 0.$$

We have seen that Property II is satisfied. But it is easy to see that $I_{KN}^{\alpha}(P:Q)$ does not satisfy Properties III and IV. We shall now find the minimum value of $I_{KN}^{\alpha}(P:Q)$ for fixed P .

From (2.64) we can see that for $p_j = q_j = 1$ and $p_i = q_i = 0$ for $i \neq j$, then both the terms in $I_{KN}^{\alpha}(P:Q)$ vanish and hence $I_{KN}^{\alpha}(P:Q) = 0$. Therefore $I_{KN}^{\alpha}(P:Q)$ satisfies Property V.

$I_{KN}^{\alpha}(P:Q)$ does not satisfy property VI. For consider the following

Example 2.3.6.1 : Let $P = (\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$ and $Q = (\frac{1}{2}, 0, \frac{1}{2})$ and let

$\alpha = \frac{1}{2}$. Here $q_2 = 0$ whereas $p_2 = \frac{1}{3} \neq 0$. Then we have

$I_{KN}^{1/2}(P:Q) = 2.0535029$, which shows that property VI is not satisfied by $I_{KN}^{\alpha}(P:Q)$.

Now we shall obtain the minimum value of $I_{KN}^{\alpha}(P:Q)$. We have

$$I_{KN}^{\alpha}(P:Q) = \frac{1}{\alpha-1} \left\{ 1 - \frac{1}{n \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}} \right\} + \frac{n^{1-\alpha}}{1-\alpha} \left\{ 1 - \frac{1}{n \sum_{i=1}^n p_i^{\alpha}} \right\}$$

In this expression both the terms are positive. We consider two cases

Case (i) : Let $\alpha > 1$ then $(\alpha-1) > 0$ and by Renyi's inequality Kapur [20] we have

$$\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} > 1$$

or
$$\frac{1}{n \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}} > 1$$

and hence

$$\frac{1}{\alpha-1} \left\{ 1 - \frac{1}{n \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}} \right\} > 0$$

and for $\alpha > 1$, $(1-\alpha) < 0$, $\sum_{i=1}^n p_i^{\alpha} < \sum_{i=1}^n p_i = 1 \implies \frac{1}{n \sum_{i=1}^n p_i^{\alpha}} > 1$ and again

$$\frac{1}{1-\alpha} \left\{ 1 - \frac{1}{n \sum_{i=1}^n p_i^{\alpha}} \right\} > 0.$$

Case (ii) : $\alpha < 1$, the same inequalities may be obtained.

Since $I_{KN}^{\alpha}(P:Q)$ is a sum of two positive terms, $\min_Q(P:Q)$ is also

sum of the minimums of the two terms. But we know by its positivity the minimum of the 1st term for a fixed $\{p_i\}$ is zero, which is attained when $q_i = p_i \forall i = 1, \dots, n$. The second term

does not involve Q and hence it remains as it is. Thus we conclude our results in

Proposition 2.3.6.4 : The minimum of $I_{KN}^{\alpha}(P:Q)$ for a fixed $\{p_i\}$ is attained when $q_i = p_i \forall i = 1, \dots, n$ and is equal to

$$\frac{n^{1-\alpha}}{1-\alpha} \left\{ 1 - \frac{1}{\sum_{i=1}^n p_i^{\alpha}} \right\}.$$

$I_{KN}^{\alpha}(P:Q) = \frac{n^{1-\alpha}}{1-\alpha}$ for $p_i = q_i = \frac{1}{n} \forall i$, but we do not know if this constitutes a minimax point as we don't know if $\frac{n^{1-\alpha}-1}{1-\alpha}$ is

$$\text{maximum for } \frac{n^{1-\alpha}}{1-\alpha} \left\{ 1 - \frac{1}{\sum_{i=1}^n p_i^{\alpha}} \right\}.$$

We are also unable to either prove or disprove $I_{KN}^{\alpha}(P:Q)$ is a convex function of q_i 's. $I_{KN}^{\alpha}(P:Q)$ is neither additive nor is subadditive. We can easily give examples proving our claim with this we complete discussion of $I_{KN}^{\alpha}(P:Q)$ and its properties.

2.3.6.4 Van der Lubbe's Second Measure of Inaccuracy :

It is defined as follows :

$$I_{L2}^{\alpha}(P:Q) = \frac{1}{\alpha-1} \left\{ 1 - \frac{\sum p_i^{\alpha}}{\sum p_i^{\alpha} q_i^{1-\alpha}} \right\} \quad (2.66)$$

We shall consider $I_{L2}^{\alpha}(U:U) = \frac{1}{\alpha-1} \{ 1 - n^{1-\alpha} \} = f(n)$ (say).

We then have $f'(n) = n^{-\alpha} > 0$. Therefore we have

Proposition 2.3.6.5 : $I_{L2}^{\alpha}(U:U)$ is a monoconically increasing function of n .

so the inaccuracy measure $I_{L2}^{\alpha}(P:Q)$ satisfies Properties I and II. However it can be easily seen that it does not satisfy properties III and IV.

For $p_j = q_j = 1$ for some j and $p_i = q_i = 0$ for $i \neq j$, $I_{L2}^{\alpha}(P:Q) = 0$

Before going in to discuss any other properties, we would like to discuss the convexity property of $I_{L2}^{\alpha}(P:Q)$ with respect to q_i 's.

Proposition 2.3.6.6 : $I_{L2}^{\alpha}(p_1, \dots, p_n; q_1, \dots, q_n)$ is convex with respect to q_i 's if and only if $0 < \alpha < 1$.

Proof : We have

$$I_{L2}^{\alpha}(p_1, \dots, p_n; q_1, \dots, q_n) = \frac{1}{\alpha-1} \left\{ 1 - \frac{\sum p_i^{\alpha}}{\sum p_i q_i^{1-\alpha}} \right\}$$

Now consider

$$f(x_1) = \frac{1}{\alpha-1} \frac{C}{A+B+Cx_1^{1-\alpha}}, \quad 0 \leq x_1 \leq 1 \quad (2.67)$$

then we have

$$f'(x_1) = \frac{C x_1^{-\alpha}}{(A+B+Cx_1^{1-\alpha})^2} \quad \text{and}$$

$$f''(x_1) = \frac{-C^2 (A+B+Cx_1^{1-\alpha})^{-\alpha-1} - 2x_1^{-\alpha} C^2 (1-\alpha) x_1^{-\alpha}}{(A+B+Cx_1^{1-\alpha})^3},$$

Now if we have $(A+B+Cx_1^{1-\alpha}) \geq 0$ then we get for $0 < \alpha < 1$

$$f''(x_1) \leq 0.$$

That is $f(x_1)$ is concave. Then, $\sum_{i=1}^n f(x_i)$ is also concave, but

$\frac{1}{\alpha-1} - n \sum_{i=1}^n f(x_i) = I_{K2}^{\alpha}(p_1, \dots, p_n; q_1, \dots, q_n)$ if we take $C = p_i^{\alpha}$

and $A = \sum_{k=1}^{i-1} p_k^{\alpha} q_k^{1-\alpha}$ and $B = \sum_{k=i-1}^n p_k^{\alpha} q_k^{1-\alpha}$, is convex. That completes the proof because $A+B+Cx_i^{1-\alpha}$ is always positive.

Now we know that Lagrange's multipliers method gives the minimum of $I_{L2}^{\alpha}(P:Q)$. So we shall apply it and find the minimum of $I_{L2}^{\alpha}(P:Q)$ for fixed $\{p_i\}$.

Proposition 2.3.6.7 : For fixed $\{p_i\}$, $I_{L2}^{\alpha}(P:Q)$ is minimum when $q_i = p_i$, $\forall i = 1, \dots, n$, and the minimum value is obtained as

$$\min_Q I_{L2}^{\alpha}(P:Q) = \frac{1}{\alpha-1} \{1 - p_i^{\alpha}\}$$

for $0 < \alpha < 1$.

Proof : Let

$$L \equiv \frac{1}{\alpha-1} \left\{ 1 - \frac{\sum_{i=1}^n p_i^{\alpha}}{\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha}} \right\} - \left\{ \sum_{i=1}^n q_i - 1 \right\}.$$

On equating $\frac{\partial L}{\partial q_i}$ to zero and solving for q_i we obtain

$$q_i = p_i \left\{ \frac{\sum_{i=1}^n p_i^{\alpha}}{n \left(\sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} \right)^2} \right\}^{\frac{1}{\alpha}} \quad (2.68)$$

where λ is to be eliminated from the equation $\sum_{i=1}^n q_i = 1$. We get

$$\lambda^{-1/\alpha} \left\{ \frac{\sum p_i^{\alpha}}{\left(\sum p_i^{\alpha} q_i^{1-\alpha} \right)^2} \right\}^{-\frac{1}{\alpha}}. \quad (2.69)$$

Therefore we have from (2.68) and (2.69)

$$q_i = p_i \quad \forall i = 1, \dots, n.$$

This distribution $q_i = p_i$ gives the minimum of $I_{L2}^\alpha(P:Q)$ for only $0 < \alpha < 1$, because $I_{L2}^\alpha(P:Q)$ is convex only for $0 < \alpha \leq 1$. We can not conclude any thing for $\alpha > 1$. By substituting $q_i = p_i \quad \forall i = 1, \dots, n$ in (2.66) we get the minimum value of $I_{L2}^\alpha(P:Q)$ as has been stated.

The minimum of this inaccuracy measure turned out be the Havrda-Charvats' measure of information, albeit for $0 < \alpha \leq 1$. Therefore we know that, for $0 < \alpha < 1$, $p_i = q_i = \frac{1}{n}$ gives a minimax point for $I_{L2}^\alpha(P:Q)$, because from proposition 2.3.6.7 we have $p_i = q_i$ giving us the minimum of $I_{L2}^\alpha(P:Q)$ and $p_i = \frac{1}{n}$ giving us the maximum of this minimum, $\frac{n^{1-\alpha}-1}{1-\alpha}$, which tends to $\ln n$ as α approaches 1. So this measure satisfies Properties V, VII and VIII the last only for $0 < \alpha \leq 1$.

$I_{L2}^\alpha(P:Q)$ does not satisfy property VI as it has a finite value even if $p_i \neq 0$ for some i whereas $q_i = 0$. But if this is true for every $i = 1, \dots, n$, then $I_{L2}^\alpha(P:Q)$ tends to infinity. Therefore we say that $I_{L2}^\alpha(P:Q)$ satisfies Property VI in only a particular case, but not in general.

One can easily verify that this measure is neither additive nor subadditive.

We had already discussed the convexity of $I_{L2}^\alpha(P:Q)$.

In the following table the following notation as followed

- Yes - the property is satisfied
 No - the property is not satisfied
 Yes/No - the property is conditionally satisfied
 - - Unknown.

2.3.7 Measures of Inaccuracy and their Properties :

Properties Measures	1	2	3	4	5	6	7	8	9	10	11
1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	Yes	Yes	No	Yes	Yes	No	No	No	Yes	No	Yes
3	Yes	Yes	No	Yes	Yes	No	No	No	No	No	Yes
4	Yes	Yes/No	No	No	Yes	Yes	No	No	No	Yes	No
5	Yes	Yes/No	No	No	Yes	-	No	No	No	Yes	No
6	Yes	Yes/No	No	No	-	-	No	No	No	Yes	No
7	Very Generalized Measure. We consider only special Cases.										
8	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	Yes/No
9	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	No	Yes/No
10	Yes/No	Yes	No	No	Yes	No	Yes	No	No	No	-
11	Yes	Yes	No	No	Yes	Yes/No	Yes	Yes	No	No	Yes/No

Table 2.3.7

with that we conclude Chapter 2.

Chapter 3

Independence Inequality and Subadditivity for Measures of Directed Divergence

3.1 Introduction : A.B. El-Sayeed had introduced in 1977 [5] an inequality for measures of entropy which he called Independence inequality.

Let $P*Q$ denote a joint probability distribution with P and Q being the marginal prob. distn. and let $H(.)$ denote any entropy measure. Then the inequality

$$H(P*Q) \leq H(PQ) \quad (3.1)$$

where PQ is the product distribution of P and Q is called the independence inequality, and the entropy measure $H(.)$ is said to be satisfy the Independence Inequality (hence forth denoted by 'the I.I.') for that particular set of prob. distn.

Here we note that the I.I. is very much a property of the distributions as well as that of measures of entropy.

El-Sayeed had obtained several types of probability distributions for which the I.I. is satisfied by $H_\alpha, H^\alpha, H_{\alpha,\beta}$ and $H^{\alpha,\beta}$ [for definitions of these measures of entropy, refer Chap. 2] . He had also obtained a set of distributions for which none of the above stated measures of entropy satisfy the I.I.

The significance of the I.I. lies in that result of El-Sayeed which states that if a measure of entropy is additive

and satisfies the I.I., then it is subadditive, for that joint probability distributions. For the definitions of additivity and subadditivity, refer to Chap. 2.

In this chapter we generalize the I.I. to the measures of directed divergence and obtain various prob. distn. for which $D_\alpha, D^\alpha, D_{\alpha,\beta}$ and $D^{\alpha,\beta}$ satisfy the I.I. We also obtain a prob. distn. for which the I.I. is satisfied by none of these measures of directed divergence. We do all this in the following section.

3.2 The I.I. for measures of Directed Divergence :

Let $P*Q$ and $R*S$ be any two joint prob. distns. with P, Q and R, S being their marginal prob. distns. respectively. Let $P \cdot Q = (\pi_{jk})_{j,k=1,\dots,n}$ and let $R \cdot S = (w_{jk})_{j,k=1,\dots,n}$. Let PQ and RS denote the product distributions of P and Q and R and S respectively. Now if $D(\cdot;\cdot)$ is any directed divergence measure, we define the following :

Sub Additivity : $D(P*Q;R*S)$ is said to be subadditive for $P*Q$ and $R*S$ if

$$D(P*Q : R*S) \leq D(P:R) + D(Q:S) \quad (3.2)$$

and if the inequality (3.2) is satisfied for all $P \cdot Q$ and $R \cdot S$ then D is said to be subadditive.

Additivity : $D(PQ:RS)$ is said to be additive for PQ and RS if

$$D(PQ:RS) = D(P:R) + D(Q:S) \quad (3.3)$$

and if the equality (3.3) holds for all independent prob. distns. then D is said to be additive.

Independence Inequality : If $D(P*Q:R\ S)$ satisfies

$$R(P*Q;R*S) \leq D(PQ:RS) \quad (3.4)$$

then D is said to satisfy the I.I. for $P\ Q$ and $R\ S$. If the inequality (3.4) is satisfied for all joint prob. distns. $P\ Q$ and $R\ S$ then D is said to satisfy the I.I.

Only the Kullback-Leibler measure of directed divergence satisfies all the three concepts defined above. The Renyi (R_α), Havrda-Charvat (D^α) Kapur ($D_{\alpha,\beta}$) and Sharma and Taneja ($D^{\alpha,\beta}$) satisfy the I.I. for certain distributions and examples are given in this chapter to show that they do not satisfy it for others.

But to begin with we establish the connection between the three concepts defined above.

Theorem 3.2.1 : (i) If D is additive and satisfies the I.I. for a set of distributions then it is subadditive for the set of distributions.

(ii) If D is subadditive and additive then it satisfies the I.I.

Proof : (i) From (3.3) we have $D(PQ:RS) = D(P:R) + D(Q:S)$ and from (3.4)

$$D(P*Q:R*S) \leq D(PQ:RS).$$

therefore we get $D(P*Q:R*S) \leq D(P:R) + D(Q:S)$ which is what is to be shown.

(ii) From (3.2) and (3.3) we have $D(P Q:R S) \leq D(P:R)+D(Q:S) = D(PQ:RS)$ which means that D satisfies the I.I. for $P*Q$ and $R*S$.

We can verify our results in the following example.

Example 3.2.1 : Let $D(P:Q) = D^\alpha(P:Q)$. Also let P be any arbitrary distribution and let Q, R and S be uniform distributions. $P*Q$ is arbitrary but we choose $R*S = RS$. At a later stage [Proposition 3.2.1] we shall show that D^α indeed satisfies the I.I. for this set of distributions. Here we shall establish that D^α is both additive and subadditive for this set of distributions.

Consider the following relation (c.f. Chap. 2) for $D^\alpha(PQ:RS)$:

$$D^\alpha(PQ:RS) = D^\alpha(P:R) + D^\alpha(Q:S) + (\alpha-1) D^\alpha(P:R) D^\alpha(Q:S) \quad (3.5)$$

Because $Q = S$, we have $D^\alpha(Q:S) = 0$.

Therefore D^α is additive for PQ and RS .

Now we shall show that D^α is subadditive for $P*Q$ and $R*S$. For, consider

$$\begin{aligned} D^\alpha(P*Q:R*S) &= \frac{1}{2^{1-\alpha}-1} \left[\sum_{j,k} \left(\frac{q_{jk}}{n} \right)^\alpha \left(\frac{1}{n^2} \right)^{\alpha-1} - 1 \right] \\ &= \frac{1}{2^{1-\alpha}-1} \left[\sum_{j,k} \frac{1}{n} q_{jk}^\alpha n^{3(1-\alpha)} - 1 \right] \end{aligned}$$

$$\leq \frac{1}{2^{1-\alpha}-1} [n^{3(1-\alpha)} \sum_j (\sum_k \frac{1}{n} q_{jk})^\alpha - 1]$$

[by convexity of $x \rightarrow x^\alpha$ for $\alpha > 1$
and $2^{1-\alpha} - 1 < 0$. If $0 \leq \alpha < 1$,
 $x \rightarrow x^\alpha$ is concave and $2^{1-\alpha} - 1 > 0$.
So in both cases the inequality is
same.]

$$= \frac{1}{2^{1-\alpha}-1} [n^{2-3} \sum_j p_j^\alpha (r_j)^{\alpha-1} - 1]$$

$$\leq \frac{1}{2^{1-\alpha}-1} [\sum_j p_j^\alpha r_j^{\alpha-1} - 1] \quad \because n^{2-3\alpha} < 1 \text{ for } \alpha < 1 \\ > 1 \text{ for } \alpha > 1$$

We have $D^\alpha(P*Q:R*S) \leq D^\alpha(P:R) + D^\alpha(Q:S)$.

Therefore we get that $D^\alpha(P:R:S)$ is subadditive. We can easily construct examples such that a directed divergence measure satisfies only the I.I. and does not satisfy the other two properties.

Lemma 3.2.1 : Let $\alpha \in \mathbb{R}$ and $\alpha \neq 1$. Then, for a given set of prob. distns. the I.I. is either satisfied by both D^α and D_α or is satisfied by neither.

Proof : $D^\alpha(P:Q) = (\sum_j p_j^\alpha q_j^{1-\alpha} - 1), \quad = (\alpha-1)^{-1} \quad (3.6)$

and $D_\alpha(P:Q) = \log (\sum_j p_j^\alpha q_j^{1-\alpha} - 1). \quad (3.7)$

The I.I. for D^α says :

$$(\sum_{j,k} p_{jk}^\alpha q_{jk}^{1-\alpha} - 1) \leq (\sum_{j,k} (p_j q_k)^\alpha (r_j s_k)^{1-\alpha} - 1)$$

$$\text{i.e.} \quad \left(\sum_{j,k} \pi_{jk}^{\alpha} w_{jk}^{1-\alpha} \right) \leq \left(\sum_{j,k} (p_j q_k)^{\alpha} (r_j s_k)^{1-\alpha} \right) \quad (3.8)$$

And for D_{α} , the I.I. is

$$\mu \log \sum_{j,k} \pi_{jk}^{\alpha} w_{jk}^{1-\alpha} \leq \mu \left(\sum_{j,k} (p_j q_k)^{\alpha} (r_j s_k)^{1-\alpha} \right) \quad (3.9)$$

Now it is clear that (3.8) and (3.9) are equivalent, because $x \rightarrow \log x$ is an increasing function. That completes the proof of Lemma 3.2.1.

So, in what follows it is enough to prove or disprove the I.I. for either D^{α} or D_{α} and the proof or the otherwise of the other follows immediately.

Proposition 3.2.1 : There exist α, β and prob. distns. such that the entropies $D^{\alpha}, D_{\alpha}, D_{\alpha, \beta}$ and $D^{\alpha, \beta}$ do not satisfy the I.I.

(For the sake of convinience, we give definitions of these entropies again : Here we consider their normalized forms

$$(i) \quad D^{\alpha}(P:Q) = \frac{1}{2^{1-\alpha}-1} \left[\sum_j p_j^{\alpha} q_j^{\alpha-1} - 1 \right] \quad \alpha \neq 1 \quad (3.10)$$

$$(ii) \quad D_{\alpha}(P:Q) = \frac{1}{2^{1-\alpha}-1} \log \sum_j p_j^{\alpha} q_j^{\alpha-1} \quad \alpha \neq 1 \quad (3.11)$$

$$(iii) \quad D_{\alpha, \beta}(P:Q) = \frac{1}{\beta-\alpha} \log \frac{\sum_j p_j^{\alpha+\beta-1} q_j^{-1+\alpha}}{\sum p_i^{\beta}} \quad \alpha \neq \beta, \\ 0 \leq \alpha < 1, \beta > 1 \\ \text{or } 0 \leq \beta < 1, \alpha > 1 \quad (3.12)$$

$$(iv) \quad D^{\alpha, \beta}(P:Q) = \frac{1}{\beta-\alpha} \left\{ \sum_j p_j^{\alpha} q_j^{-1+\alpha} - \sum_j p_j^{\beta} q_j^{-1+\beta} \right\} \quad \text{-do-} \quad (3.13))$$

Proof : (i) Let $\alpha = 2$, $P*Q = \begin{pmatrix} 0.2 & 0.3 \\ 0.25 & 0.25 \end{pmatrix}$ and $R*S = \begin{pmatrix} 0.2 & 0.1 \\ 0.3 & 0.4 \end{pmatrix}$.
Then we get $P = (0.5, 0.5)$, $Q = (0.45, 0.55)$, $R = (0.3, 0.7)$ and $S = (0.5, 0.5)$.

$$PQ = \begin{pmatrix} 0.225 & 0.275 \\ 0.225 & 0.275 \end{pmatrix} \text{ and } RS = \begin{pmatrix} 0.15 & 0.15 \\ 0.35 & 0.35 \end{pmatrix}.$$

Now we have

$$\left(\sum_{j,k} \pi_{jk}^{\alpha} \pi_{jk}^{-1+\alpha} - 1 \right) = 1.8785 \text{ and}$$

$$\left(\sum_{j,k} (p_j q_k)^{\alpha} (r_j s_k)^{-1+\alpha} - 1 \right) = 1.87375 \text{ which shows that } D^{\alpha}$$

(and hence by Lemma 3.2.1, D_{α}) does not satisfy the I.I.

(iii) Let $\alpha = 2$, $\beta = 0.5$ and $P*Q$ and $R*S$ be as in (i). Then we get

$$\frac{1}{\beta-\alpha} \log \frac{\sum_{j,k} n_{jk}^{\alpha+\beta-1} \pi_{jk}^{-1+\alpha}}{\sum_{j,k} \pi_{jk}^{\beta}} = 0.8094736$$

$$\frac{1}{\beta-\alpha} \log \left\{ \frac{\sum_{j,k} (p_j q_k)^{\alpha+\beta-1} (r_j s_k)^{1+\alpha}}{\sum_{j,k} (p_j q_k)^{\beta}} \right\} = 0.80129.$$

which proves that $D_{\alpha,\beta}$ does not satisfy the I.I. for the given set of distributions.

(iv) We have for $\alpha = 2$, $\beta = 0.5$,

$$\frac{1}{-\alpha+\beta} \left\{ \sum_{j,k} \pi_{jk}^{\alpha} \pi_{jk}^{-1+\alpha} - \sum_{j,k} \pi_{jk}^{\beta} \pi_{jk}^{-1+\beta} \right\} = 2.9164887 \text{ and}$$

$$\frac{1}{-\alpha+\beta} \left\{ \sum_{j,k} (p_j q_k)^{\alpha} (r_j s_k)^{-1+\alpha} - \sum_{j,k} (p_j q_k)^{\beta} (r_j s_k)^{-1+\beta} \right\} = 2.8025452.$$

So we have been able to find a set of distributions for which none of the four directed divergence measures $D^\alpha, D_\alpha, D_{\alpha,\beta}$, and $D^{\alpha,\beta}$ satisfies the I.I.

In the following Example, we construct a set of distributions for which D^α and hence D_α satisfy the I.I.

Example 3.2.2 : Let P Q and R S be defined by the conditional probability matrices :

$$q_{ik} = \begin{bmatrix} 1-p & \frac{p}{n-1} & \dots & \frac{p}{n-1} \\ \frac{p}{n-1} & 1-p & \dots & \frac{p}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p}{n-1} & \frac{p}{n-1} & \dots & 1-p \end{bmatrix}_{n \times n} \quad \text{and} \quad s_{ik} = \begin{bmatrix} 1-p & \frac{p}{n-1} & \dots & \frac{p}{n-1} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \frac{p}{n-1} & \frac{p}{n-1} & \dots & 1-p \end{bmatrix}_{n \times n} \quad (3.14)$$

where $0 \leq p \leq 1$, $p_i = \frac{1}{n} \forall i = 1, \dots, n$, $R = (r_1, \dots, r_n)$

$= (1-p, 0, 0, \dots, 0, p)$.

Then we have

$$\pi_{jk} = p_j q_{jk} = \frac{1}{n} q_{jk}, \quad jk=r_j s_k = \begin{bmatrix} (1-p)^2 & \frac{p(1-p)}{n-1} & \dots & \frac{p(1-p)}{n-1} \\ 0, 0, 0, & \dots & 0 \\ 0, 0, & 0, & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \frac{p^2}{n-1} & \frac{p^2}{n-1} & \dots & p(1-p) \end{bmatrix}_{n \times n}$$

Now $q_k = \sum_{j=1}^n \pi_{jk} = \frac{1}{n} \sum_{j=1}^n q_{jk} = \frac{1}{n} \forall j = 1, \dots, n$.

and finally we have $S = (s_1, s_2, \dots, s_n) = (1-p, \frac{p}{n-1}, \frac{p}{n-1}, \dots, \frac{p}{n-1})$.

We obtain after making the calculations

$$\sum_{j,k} (\pi_{jk})^\alpha (j_k)^{\alpha-1} = n^{-\alpha} [(1-p)^{\alpha-1} + p^{\alpha-1}] [(1-p)^{2\alpha-1} + (n-1)^{2(1-\alpha)} p^{2\alpha-1}] \quad (3.15)$$

and

$$\sum_{j,k} (p_j q_k)^\alpha (r_j s_k)^{\alpha-1} = n^{-2\alpha} [(1-p)^{\alpha-1} + p^{\alpha-1}] [(1-p)^{\alpha-1} + (n-1)^{2-\alpha} p^{\alpha-1}] \quad (3.16)$$

Let $f(p) = \frac{(1-p)^{2\alpha-1} + (n-1)^{2(1-\alpha)} p^{2\alpha-1}}{(1-p)^{\alpha-1} + (n-1)^{2-\alpha} p^{\alpha-1}}$ then we have (3.17)

$$f(0) = 1 \text{ and } f(1) = \left(\frac{1}{n-1}\right)^\alpha$$

$$[(1-p)^{\alpha-1} + (n-1)^{2-\alpha} p^{\alpha-1}] [-(2\alpha-1)(1-p)^{2\alpha-2} + (n-1)^{2(1-\alpha)} (2\alpha-1) p^{2\alpha-2}] - [(1-p)^{2\alpha-1} + (n-1)^{2(1-\alpha)} p^{2\alpha-1}]$$

$$f'(p) = \frac{[-(\alpha-1)(1-p)^{\alpha-2} + (n-1)^{2-\alpha}(\alpha-1)p^{\alpha-2}]}{[(1-p)^{\alpha-1} + (n-1)^{2-\alpha} p^{\alpha-1}]^2} \quad (3.19)$$

And $f'(p) = 0 \iff p = p_0 = \left(\frac{n-1}{n}\right)$ and $f(p_0) = \left(\frac{1}{n}\right)^\alpha$. (3.20)

From (3.18) and (3.20) we can see that

$$0 < \left(\frac{1}{n}\right)^\alpha = f(p_0) < \left(\frac{1}{n-1}\right)^\alpha = f(1) \quad (3.21)$$

∴ The point $p = p_0$ is a minimum for $f(p)$ i.e., we have

$$(3.17) \quad n^{+\alpha} f(p) \leq n^\alpha n^{-\alpha} = 1$$

or $[(1-p)^{2\alpha-1} + (n-1)^{2(1-\alpha)} p^{2\alpha-1}] \leq n^{-\alpha} [(1-p)^{\alpha-1} + (n-1)^{2-\alpha} p^{\alpha-1}]$

$$\text{or } \sum_{j,k} (\pi_{jk})^\alpha (w_{jk})^{\alpha-1} \leq \sum_{j,k} (p_j q_k)^\alpha (r_j s_k)^{\alpha-1}.$$

Now because $\mu = (2^{1-\alpha} - 1)^{-1} > 0$, we get the I.I. satisfied for both D^α and D_α . This proof holds for all values of α , $\alpha \neq 1$.

In the next proposition, we shall consider the I.I. for more general distributions than the ones considered in Example 3.2.2.

Proposition 3.2.2 : For arbitrary $P*Q$ with $Q = U$ and $R = S = U$, where U is the uniform distributions with n components and $R*S = U_{n \times n}$, the following directed divergences satisfy the I.I.

- (i) D^α , $\alpha > 0$, $\alpha \neq 1$
- (ii) D_α , $\alpha \geq 0$, $\alpha \neq 1$
- (iii) $D_{\alpha,\beta}$, $\alpha+\beta-1 > 1$, $\beta < 1$ (or $\alpha+\beta-1 < 1$, $\beta > 1$)
- (iv) $D^{\alpha,\beta}$, $\alpha > 1$, $\beta < 1$ or $\alpha < 1$, $\beta > 1$.

Proof : (i) Let $0 \leq \alpha < 1$. Then we get $\mu = (2^{1-\alpha} - 1)^{-1} > 0$. We also have by concavity of $x \rightarrow x^\alpha$,

$$\begin{aligned} \sum_k \frac{1}{n} q_{jk}^\alpha (r_j s_{jk})^{\alpha-1} &= n^{2(1-\alpha)} \sum_k \frac{1}{n} q_{jk}^\alpha \\ &\leq n^{2(1-\alpha)} \left(\sum_k \frac{1}{n} q_{jk} \right)^\alpha \\ &= n^{2(1-\alpha)} \frac{1}{n} \end{aligned}$$

$$\text{or } \sum_k q_{jk}^\alpha (r_j s_{jk})^{\alpha-1} \leq n^{2(1-\alpha)} n^{1-\alpha} = n^{2(1-\alpha)} \sum_k q_k^\alpha, \forall j=1,2,\dots,n$$

(3.22)

Now multiplying both sides of (3.22) by p_j^α and summing over j , we get

$$\begin{aligned} \sum_{j,k} (p_j q_{jk})^\alpha (r_j s_{jk})^{\alpha-1} &\leq \sum_{j,k} (p_j q_k)^\alpha n^{2(1-\alpha)} \\ &= \sum_{j,k} (p_j q_k)^\alpha (r_j s_k)^{\alpha-1} \end{aligned} \quad (3.23)$$

From (3.23) and because $\mu > 0$ we get

$$\sum_{j,k} (p_j q_{jk})^\alpha (r_j s_{jk})^{\alpha-1} \leq \sum_{j,k} (p_j q_k)^\alpha (r_j s_k)^{\alpha-1} \quad (3.24)$$

(3.24) implies the I.I. for D^α .

(b) $\alpha > 1$. Now $\mu < 0$ and $x \rightarrow x^\alpha$ is convex, so the inequality in (3.23) is reversed, but $\mu < 0$, therefore we again obtain (3.24).

(ii) Using (i) and Lemma 3.2.1 we get the I.I. for D_α .

(iii) Let $\beta < 1$, $\alpha + \beta - 1 > 1$. Now $\mu = (\beta - \alpha)^{-1} < 0$ and $x \rightarrow x^{\alpha + \beta - 1}$ is convex. Therefore, we get by proceeding in the same way as we did in (i) above,

$$\sum_{j,k} (\pi_{jk})^{\alpha + \beta - 1} (w_{jk})^{-1 + \alpha} \geq \sum_{j,k} (p_j q_k)^{\alpha + \beta - 1} (r_j s_k)^{-1 + \alpha} \quad (3.24)$$

and

$$\sum_{j,k} (\pi_{jk}) \leq \sum_{j,k} (p_j q_k) \quad (3.25)$$

From (3.24) and (3.25)

$$\log \frac{\sum_{j,k} (\pi_{jk})^{\alpha + \beta - 1} (w_{jk})^{-1 + \alpha}}{\sum_{j,k} (\pi_{jk})^\beta} \geq \log \frac{\sum_{j,k} (p_j q_k)^{\alpha + \beta - 1} (r_j s_j)^{-1 + \alpha}}{\sum_{j,k} (p_j q_k)^\beta} \quad (3.26)$$

and because $\mu < 0$,

$$\log \left\{ \frac{\sum_{j,k} (\pi_{jk})^{\alpha+\beta-1} (w_{jk})^{-1+\alpha}}{\sum_{j,k} (\pi_{jk})^{\beta}} \right\} \leq \log \left\{ \frac{\sum_{j,k} (p_{jq_k})^{\alpha+\beta-1} (r_{js_k})^{-1+\alpha}}{\sum_{j,k} (p_{jq_k})^{\beta}} \right\} \quad (3.27)$$

The last inequality is what we intended to prove.

Now if we consider $\beta > 1$ and $\alpha+\beta-1 < 1$, then $\mu > 0$, $x \rightarrow x^{\alpha+\beta-1}$ is concave. Therefore the inequality (3.26) is reversed and once again we obtain (3.27).

$$(iv) \quad D^{\alpha,\beta}(P:Q) = \frac{1}{(-\alpha+\beta)} \left\{ \sum_j p_j^{\alpha} q_j^{-1+\alpha} - \sum_j p_j^{\beta} q_j^{-1+\beta} \right\}.$$

Now consider $\alpha > 1$, $\beta < 1$. Then $\mu = (\beta-\alpha)^{-1} < 0$. $x \rightarrow x^{\alpha}$ is convex and $x \rightarrow x^{\beta}$ is concave. Therefore we obtain the following inequalities :

$$\sum_{j,k} (\pi_{jk})^{\alpha} (w_{jk})^{-1+\alpha} \geq \sum_{j,k} (p_{jq_k})^{\alpha} (r_{js_k})^{\alpha-1} \quad (3.28)$$

$$\sum_{j,k} (\pi_{jk})^{\beta} (w_{jk})^{\beta-1} \leq \sum_{j,k} (p_{jq_k})^{\beta} (r_{js_k})^{\beta-1} \quad (3.29)$$

Now from (3.28) and (3.29) we get

$$\begin{aligned} & \left(\sum_{j,k} (\pi_{jk})^{\alpha} (w_{jk})^{\alpha-1} - \sum_{j,k} (\pi_{jk})^{\beta} (w_{jk})^{\beta-1} \right) \geq \\ & \left(\sum_{j,k} (p_{jq_k})^{\alpha} (r_{js_k})^{\alpha-1} - \sum_{j,k} (p_{jq_k})^{\beta} (r_{js_k})^{\beta-1} \right) \end{aligned} \quad (3.30)$$

But $\mu < 0$, thereby

$$\mu \left(\sum_{j,k} (\pi_{jk})^\alpha (w_{jk})^{\alpha-1} - \sum_{j,k} (\pi_{jk})^\beta (w_{jk})^{\beta-1} \right) \\ \leq \mu \left(\sum_{j,k} (p_{jq_k})^\alpha (r_{js_k})^{\alpha-1} - \sum_{j,k} (p_{jq_k})^\beta (r_{js_k})^{\beta-1} \right) \quad (3.31)$$

(3.31) is the I.I. for $D^{\alpha,\beta}$. If we consider $\alpha < 1$, $\beta > 1$, $\mu > 0$, inequality in (3.30) is reversed giving (3.31) again. That completes the proof proposition 3.2.2.

Remark : We can easily verify from the proof of Proposition 3.2.2 that the following directed divergences do not satisfy the I.I. for the same set of distributions as in the proposition :

- (i) D^α , $\alpha \leq 0$
- (ii) D_α , $\alpha \leq 0$
- (iii) $D_{\alpha,\beta}$, $\alpha+\beta-1 > 1$, $\beta < 1$ or $\alpha+\beta-1 > 1$, $\beta > 1$
- (iv) $D^{\alpha,\beta}$, $\alpha > 1$, $\beta > 1$ or $\alpha < 1$, $\beta < 1$.

Proposition 3.2.3 : If the rows of the conditional probability matrix (q_{jk}) are permutations of the same n -numbers and $R \neq S$, R and S as in Proposition 3.2.2, the following directed divergences satisfy the I.I. :

- (i) D^α , $\alpha \geq 0$, $\alpha \neq 1$
- (ii) D_α , $\alpha \geq 0$, $\alpha \neq 1$
- (iii) $D_{\alpha,\beta}$, $\alpha+\beta-1 > 1$, $\beta < 1$ or $\alpha+\beta-1 < 1$, $\beta > 1$.
- (iv) $D^{\alpha,\beta}$, $\alpha > 1$, $\beta < 1$ or $\alpha < 1$, $\beta > 1$.

Proof : Let such set of n numbers be (a_1, a_2, \dots, a_n) and

let $\sum_{i=1}^n a_i^\alpha = A$. Then

$$\sum_k q_{jk}^\alpha (r_j s_{jk})^{\alpha-1} = \sum_k a_k^\alpha n^{2(1-\alpha)} = n^{2(1-\alpha)} A \quad (3.32)$$

$$\sum_k \sum_j p_j q_{jk}^\alpha (r_j s_{jk})^{\alpha-1} = \sum_j p_j A n^{2(1-\alpha)} = n^{2(1-\alpha)} A. \quad (3.33)$$

From (3.32) and (3.33) we get

$$\sum_k q_{jk}^\alpha (r_j s_{jk})^{\alpha-1} = \sum_k \sum_j p_j q_{jk}^\alpha (r_j s_{jk})^{\alpha-1} \quad (3.34)$$

Now if $0 \leq \alpha < 1$, $\sum_k \sum_j p_j q_{jk}^\alpha (r_j s_{jk})^{\alpha-1}$

$$\leq \sum_k (\sum_j p_j q_{jk})^\alpha n^{2(1-\alpha)} = \sum_k q_k^\alpha n^{2(1-\alpha)} \quad (3.35)$$

From (3.34) and (3.35) we get

$$\sum_k q_{jk}^\alpha (r_j s_{jk})^{\alpha-1} \leq n^{2(1-\alpha)} \sum_k q_k^\alpha, \quad \forall j=1, 2, \dots, n. \quad (3.36)$$

Now multiplying with p_j^α and summing over j , we get from (3.36)

$$\sum_{j,k} (p_j q_{jk})^\alpha (r_j s_{jk})^{\alpha-1} \leq \sum_{j,k} (p_j q_k)^\alpha (r_j s_k)^{\alpha-1} \quad (3.37)$$

For $0 \leq \alpha < 1$, $\mu = (2^{1-\alpha} - 1)^{-1} > 0$

$$\mu \left(\sum_{j,k} (p_j q_{jk})^\alpha (r_j s_{jk})^{\alpha-1} - 1 \right) \leq \mu \left(\sum_{j,k} (p_j q_k)^\alpha (r_j s_k)^{\alpha-1} - 1 \right) \quad (3.38)$$

which is the I.I. for D^α (and hence for D_α). We can prove the result in the same way for $\alpha > 1$. And by taking the values of α and β in the range prescribed, we can easily obtain the result both for $D_{\alpha,\beta}$ and $D^{\alpha,\beta}$. That completes the proof of Proposition 3.2.3.

Remark : The following directed divergences do not satisfy the I.I. for distributions as in Proposition 3.2.3 :

$$(i) \quad D^\alpha, \alpha \leq 0$$

$$(ii) \quad D_\alpha, \alpha \leq 0$$

$$(iii) \quad D_{\alpha,\beta}, \alpha+\beta-1 > 1, \beta > 1 \text{ or } \alpha+\beta-1 < 1, \beta < 1$$

$$(iv) \quad D^{\alpha,\beta}, \alpha > 1, \beta > 1 \text{ or } \alpha < 1, \beta < 1.$$

Proposition 3.2.4 : Let $\alpha \geq 0$ and $\alpha \neq 1$. If the elements of the conditional probability matrix (q_{jk}) satisfy the equation $\sum_k q_{jk}^\alpha = A$ for all $j = 1, 2, \dots, n$ where A is some constant and R, S, R and S as in Proposition 3.2.2 then D^α and D_α satisfy the I.I.

Proof : Note that in the proof of Proposition 3.2.3, we used the row permutation property of (q_{jk}) in obtaining the equation :

$$\sum_k q_{jk}^\alpha (r_j s_{jk})^{\alpha-1} = A n^{2(1-\alpha)}, \text{ a constant} \quad (3.39)$$

Therefore all the proof of (i) and (ii) of Proposition 3.2.3 remains valid for all distributions satisfying (3.39), not necessarily the row permutation property of (q_{jk}) . That completes the proof of Proposition 3.2.4.

Remark : (i) The I.I. is satisfied by the directed divergences noted in Proposition 3.2.2 if $P = U$, where U is the uniform distribution and R, S, R and S exactly as in Proposition 3.2.2.

$$\begin{aligned}
 \underline{\text{Proof}} : L &:= \sum_{j,k} (\pi_{jk})^\alpha (w_{jk})^{\alpha-1} = \sum_{j,k} (p_j q_{jk})^\alpha (r_j s_{jk})^{\alpha-1} \\
 &= \sum_{j,k} \left(\frac{q_{jk}}{n}\right)^\alpha \left(\frac{1}{n}\right)^{\alpha-1} \\
 &= n^{2(1-\alpha)} \left(\frac{1}{n}\right)^\alpha \sum_{j,k} (q_{jk})^\alpha \quad (3.40)
 \end{aligned}$$

$$\begin{aligned}
 R &:= \sum_{j,k} (p_j q_k)^\alpha (r_j s_k)^{\alpha-1} = n^{2(1-\alpha)} \left(\frac{1}{n}\right)^\alpha n \sum_k q_k^\alpha \\
 &= n^{2(1-\alpha)} \left(\frac{1}{n}\right)^\alpha n \sum_k \left(\sum_j p_j q_{jk}\right)^\alpha \\
 &= n^{2(1-\alpha)} \left(\frac{1}{n}\right)^\alpha n^{1-\alpha} \sum_k \left(\sum_j q_{jk}\right)^\alpha \quad (3.41)
 \end{aligned}$$

Now if we take $0 \leq \alpha \leq 1$, we have by concavity of $x \rightarrow x^\alpha$,

$$\sum_j \frac{1}{n} q_{jk}^\alpha \leq \left(\sum_j \frac{1}{n} q_{jk}\right)^\alpha = n^{1-\alpha} \left(\sum_j q_{jk}\right)^\alpha$$

$$\therefore \sum_{j,k} q_{jk}^\alpha \leq n^{1-\alpha} \sum_k \left(\sum_j q_{jk}\right)^\alpha. \quad (3.42)$$

From (3.40), (3.41) and (3.42) we get $L \leq R$, and because

$\mu = (2^{1-\alpha} - 1)^{-1} > 0$ the I.I. is proved D^α and D_α .

For $\alpha > 1$, we get $L \geq R$ and $\mu < 0$, thereby proving the I.I. for

D^α and D_α again. For $D_{\alpha,\beta}$ and $D^{\alpha,\beta}$ the proof is similar to that of Proposition 3.2.2.

Corollary 3.2.1 : The I.I. is satisfied by the directed divergences noted in Proposition 3.2.2 if the joint probability distribution $P*Q$ is uniform and $R*S$ as in the Proposition 3.2.2.

Proof : If $P*Q$ is uniform, then

$$p_j = \sum_k \pi_{jk} = \sum_{k=1}^n \frac{1}{n^2} = n \times \frac{1}{n^2} = \frac{1}{n} \quad \forall j = 1, \dots, n.$$

\therefore The condition of the Proposition 3.2.5 is satisfied.

Note here that if $P*Q$ is uniform and because $R*S$ is also assumed to be uniform, we have both L and $R = 0$, proving the Corollary.

Corollary 3.2.2 : If the conditional probability distribution (q_{jk}) is uniform then also the I.I. is satisfied by the divergences of Proposition 3.2.2 when $R*S$ is uniform.

Proof : If the conditional probability distribution is uniform, then it satisfies the row permutation property. We also have

$$q_k = \sum_j \pi_{jk} = \sum_j p_j q_{jk} = \sum_j p_j \frac{1}{n} = \frac{1}{n}.$$

Therefore this corollary can be considered as a corollary of Proposition 3.2.2 as well.

Now we shall summarize the cases we have considered in Propositions 3.2.2 - 3.2.5 in the following theorem.

Theorem 3.2.2 : The I.I. is satisfied by the following directed divergences

- (i) $D^\alpha, \alpha \geq 0, \alpha \neq 1$
- (ii) $D_\alpha, \alpha \geq 0, \alpha \neq 1$
- (iii) $D_{\alpha, \beta}, \alpha + \beta - 1 < 1, \beta > 1$ or $\alpha + \beta - 1 > 1, \beta < 1$
- (iv) $D^{\alpha, \beta}, \alpha > 1, \beta < 1$ or $\alpha < 1, \beta > 1$

if any of the following conditions is satisfied :

1. The probability distribution P is uniform and R*S is uniform distribution.
2. The probability distribution Q is uniform and R*S is uniform distribution.
3. The conditional probability matrix $(q_{jk})_{n \times n}$ has the row permutation property and R*S is uniform distribution.
4. If there exists a constant A s.t. $\sum_k q_{jk}^\alpha = A \forall j = 1, \dots, n$ and R*S is uniform distribution, then only (i) and (ii) satisfy the I.I.

Corollary 3.2.3 : The additive directed divergences

- (i) $D_{\alpha, \beta}, \alpha + \beta - 1 < 1, \beta > 1$ or $\alpha + \beta - 1 > 1, \beta < 1$
and
- (ii) $D_\alpha, \alpha \geq 0, \alpha \neq 1$

are subadditive for any of the probability distributions in Theorem 3.2.2.

Proof : It is a direct application of Theorem 1.

with that we conclude this chapter.

Chapter 4

Subadditivity, Superadditivity and Measures of Dependence

4.1 Introduction : Let X and Y be two random variables and let

$$\Pr(X = x_i) = p_i, \quad i = 1, 2, \dots, m \quad (4.1)$$

$$\Pr(Y = y_j) = q_j, \quad j = 1, 2, \dots, n \quad (4.2)$$

$$\Pr(X = x_i, Y = y_j) = p_{ij}, \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n \quad (4.3)$$

$$\text{and let } P = (p_1, p_2, \dots, p_m), \quad Q = (q_1, q_2, q_2, \dots, q_n) \quad (4.4)$$

$$P*Q = (p_{11}, p_{12}, \dots, p_{1n}, \dots, \dots, \dots, \dots, p_{mn})$$

$$PQ = (p_1 q_1, p_1 q_2, \dots, p_1 q_n, \dots, \dots, \dots, \dots, p_m q_n) \quad (4.5)$$

denote the corresponding probability distributions. Now

$$\sum_{j=1}^n p_{ij} = p_i, \quad i = 1, 2, \dots, m; \quad \sum_{i=1}^m p_{ij} = q_j, \quad j = 1, 2, \dots, n \quad (4.6)$$

so that for a given bivariate probability distribution $P*Q$, we can find the marginal probability distributions P and Q .

Now let $E(P)$, $E(Q)$, $E(P*Q)$ and $E(PQ)$ denote the entropies of the corresponding probability distributions according to any measure of entropy E we may use.

(i) The measure is said to be subadditive for $P*Q$ if

$$E(P*Q) \leq E(P) + E(Q) \quad (4.7)$$

i.e., if the information given by $P*Q$ is less than or equal to the sum of the information given by P and Q separately.

(ii) The measure E is said to be superadditive for $P*Q$ if

$$E(P*Q) \geq E(P) + E(Q) \quad (4.8)$$

i.e., if the information given by $P*Q$ is greater than or equal to the sum of the information given by P and Q separately.

(iii) The measure E is said to be additive for PQ if

$$E(PQ) = E(P) + E(Q) \quad (4.9)$$

i.e., the information given by PQ is equal to the sum of the informations given by P and Q separately.

We make the following remarks on the definitions :

a) Subadditivity, Superadditivity and additivity for E are defined w.r.to each bivariate probability distribution, $P*Q$ so that a measure of entropy may be subadditive (superadditive, additive) for some distribution and may not be so for others. If a measure is subadditive (superadditive, additive) for all bivariate probability distributions, it is simply said to be subadditive (superadditive, additive).

Then Shannon's measure of entropy

$$S(P) = - \sum_{i=1}^m p_i \ln p_i \quad (4.11)$$

is shown to be additive and subadditive for all distributions. As such we shall say that Shannon's measure of entropy has both additivity and subadditivity properties.

Again we shall show in the present chapter that Renyi's measure of entropy

$$R(P) = \frac{1}{1-\alpha} \ln \sum_{i=1}^n p_i^\alpha \quad \alpha \neq 1, \alpha > 0 \quad (4.12)$$

satisfies the subadditivity property for prob. distns. when $\alpha < 1$. Then every member of Renyi's family of measures of entropy for which $0 < \alpha < 1$ is subadditive. Similarly it is known that every member of Renyi's family of measures (whether $0 \leq \alpha < 1$ or $\alpha > 1$) is additive.

It is also known that every member of Havrda-Charvat's family of measures

$$H(P) = \frac{1}{1-\alpha} \left[\sum_{i=1}^n p_i^\alpha - 1 \right] \quad \alpha > 0, \alpha \neq 1 \quad (4.13)$$

is subadditive for all $\alpha > 1$, if only product of two independent distributions is considered as $P*Q$. We expect that the measures belonging to the $H(P)$ are subadditive for all $\alpha > 1$ for all distributions, but we are unable to prove it. Neither were we able to disprove it by a counter-example.

However when $\alpha > 1$, Renyi's measures of entropy are not subadditive. El-Sayeed [5] gives a numerical example of prob. distn. for which subadditivity condition is violated when $\alpha = 2$. However his example only shows that Renyi's measure of entropy for $\alpha = 2$

is not subadditive. Renyi's is a family of measures, one measure corresponding to every value of α . Different members of the family may have different properties. Then Renyi's measure not being subadditive for $\alpha = 2$ does not mean that it will not be subadditive for all values of α . In fact we know that Renyi's measure is subadditive for the limiting case, when $\alpha \rightarrow 1$. From continuity considerations we can reasonably expect that Renyi's measure will also be subadditive for all probability distributions in the neighbourhood of the parametric value unity. We investigate whether this is true and discuss similar questions for superadditivity and additivity.

Similarly for Havrda-Charvats' measure represents a family of entropy measures for different values of α and some of them may be subadditive for some parametric values and not for other. We also investigate this problem.

Again for the independent variates

$$P*Q = PQ \quad (4.14)$$

so that $E(P*Q) = E(PQ)$. The independence inequality [Chapter 3, (3.1)] requires that for random variates which are not independent, $E(P*Q) \leq E(PQ)$. If the condition is satisfied we can use $[E(PQ) - E(P*Q)]$ as a measure of the dependence of the variates P and Q .

From (4.12) and (4.13) we get

$$R(P) = \frac{1}{1-\alpha} \ln \{ (1-\alpha) H(P) + 1 \} \quad (4.15)$$

$$R(P)+R(Q)-R(P*Q) = \frac{1}{1-\alpha} \ln \frac{(1-\alpha)^2 H(P)H(Q) - (1-\alpha)(H(P)+H(Q)) + 1}{(1-\alpha)H(P*Q) + 1} \quad (4.16)$$

$$= \frac{1}{1-\alpha} \ln \frac{N}{D} \quad (4.17)$$

$$\text{where } N = (1-\alpha)^2 H(P) H(Q) - (1-\alpha)(H(P) + H(Q)) + 1$$

$$D = (1-\alpha) H(P*Q) + 1 \quad (4.18)$$

we use (4.16) and (4.17) in the next section to investigate the relationship between the subadditivity and superadditivity of $H(P)$ and $R(P)$.

4.2 On the relationship between Subadditivity and Superadditivity of Renyi's and Havrda-Charvats' Measures of Entropy

Theorem 4.2.1 : If $0 < \alpha < 1$

- a) The subadditivity of $H \implies$ the subadditivity of R
- b) The superadditivity of $R \implies$ the superadditivity of H

Proof : a) Subadditivity of H for $P*Q \implies H(P)+H(Q) \geq H(P*Q)$
 $\implies (1-\alpha)(H(P)+H(Q)) \geq (1-\alpha)H(P*Q)$ since $0 < \alpha < 1$
 $\implies 1+(1-\alpha)(H(P)+H(Q)) > 1+(1-\alpha)H(P*Q)$
 $\implies (1-\alpha)^2 H(P)H(Q) + (1-\alpha)(H(P)+H(Q)) + 1$
 $\geq (1-\alpha)H(P*Q) + 1$
 $\implies N \geq D \implies \ln \frac{N}{D} \geq 0 \implies \frac{1}{1-\alpha} \ln \frac{N}{D} \geq 0$
 $\implies R(P) + R(Q) - R(P*Q) \geq 0$
 \implies Subadditivity of R for $P*Q$.

b) Superadditivity of R for $P*Q \Rightarrow R(P) + R(Q) \leq R(P*Q)$

$$\Rightarrow \frac{1}{1-\alpha} \ln \frac{N}{D} \leq 0 \Rightarrow \frac{N}{D} \leq 1$$

$$\Rightarrow (1-\alpha)(H(P)+H(Q) - H(P*Q)) \leq -(1-\alpha)^2 H(P)H(Q) \leq 0$$

$$\Rightarrow H(P) + H(Q) - H(P*Q) \leq 0$$

\Rightarrow superadditivity of H.

Theorem 4.2.2 : If $\alpha > 1$, then

a) Subadditivity of R \Rightarrow subadditivity of H

b) Superadditivity of H \Rightarrow superadditivity of R.

Proof : a) Subadditivity of R for $P*Q \Rightarrow R(P)+R(Q) \geq R(P*Q)$

$$\Rightarrow \frac{1}{1-\alpha} \ln \frac{N}{D} \geq 0 \Rightarrow \frac{N}{D} \leq 1 \quad \because \alpha > 1$$

$$\Rightarrow (1-\alpha)^2 H(P) H(Q) + (1-\alpha)(H(P)+H(Q)) \leq (1-\alpha)H(P*Q)$$

$$\Rightarrow (1-\alpha)\{H(P)+H(Q)-H(P*Q)\} \leq -(1-\alpha)^2 H(P)H(Q) \leq 0$$

b) Superadditivity of H for $P*Q \Rightarrow H(P)+H(Q) \leq H(P*Q)$

$$\Rightarrow (1-\alpha)(H(P)+H(Q)) \geq (1-\alpha)H(P*Q)$$

$$\Rightarrow (1-\alpha)^2 H(P)H(Q) + (1-\alpha)\{H(P)+H(Q)\} + 1 \geq (1-\alpha)H(P*Q) + 1$$

$$\Rightarrow N \geq D$$

$$\Rightarrow \frac{1}{1-\alpha} \ln \frac{N}{D} \leq 0 \Rightarrow R(P)+R(Q) \leq R(P*Q)$$

\Rightarrow Superadditivity of R.

Theorem 4.2.3 :

- a) Subadditivity of H for $P*Q \implies$ Subadditivity of R for $P*Q$ if $0 < \alpha < 1$ and Superadditivity of H for $P*Q \implies$ Superadditivity of R for $P*Q$ if $\alpha > 1$.
- b) Subadditivity of R for $P*Q \implies$ subadditivity of H for $P*Q$ if $\alpha > 1$ and Superadditivity of R for $P*Q \implies$ Superadditivity of H for $P*Q$ if $0 < \alpha < 1$.

The converses of these results are not true. For, consider the following example. Theorem 4.2.3 is a restatement of Theorems 4.2.1 and 4.2.2 and hence requires no further proving.

Example 4.2.1 :

- a) Let $P*Q = \begin{pmatrix} 0.1 & 0.5 \\ 0.3 & 0.1 \end{pmatrix}$

Then for $\alpha = 0.1$, $R_\alpha(P*Q) - R_\alpha(P) - R_\alpha(Q) = -0.17759 \cdot 10^{-1} < 0$

and $\alpha = 0.1$, $H_\alpha(P*Q) - H_\alpha(P) - H_\alpha(Q) = 0.75885 > 0$.

$R_{0.1}$ is subadditive and $H_{0.1}$ is superadditive. This is a counter-example for the converse of Thm. 4.2.3.a), first statement.

- b) Again same distribution as in a) is considered. Then for $\alpha = 1.285$

$R_\alpha(P*Q) - R_\alpha(P) - R_\alpha(Q) = 0.1735 \cdot 10^{-4} > 0$, $R_{1.285}$ is superadditive.

$H_\alpha(P*Q) - H_\alpha(P) - H_\alpha(Q) = 0.039911 < 0$, $H_{1.285}$ is subadditive.

∴ This constitutes a counter-example for the converse of Thm. 4.2.3.a) second statement.

- c) Case b) above works as a counter-example for the converse of Thm. 4.2.3, b) the first statement, because there, for $\alpha = 1.285$, H_α is subadditive where as R_α is superadditive.
- d) Case a) works as a counter-example for the converse of Thm. 4.2.3, b) the second statement, because there, for $\alpha = 0.1$, H_α is superadditive and R_α is subadditive.

4.3 Subadditivity and Superadditivity of Renyi's Measure of Entropy

Theorem 4.3.1 : Renyi's Measure of entropy satisfies the inequalities

$$R(P*Q) \geq R(P) \text{ and } R(P*Q) \geq R(Q) \text{ for all } P*Q \quad (4.19)$$

Proof : Since $p_{ij} = p_i p_{j,i}$ where $p_{j,i}$ is the conditional probability that the second experiment results in the j th outcome when the outcome of the expt. is known to have been x_i , we get

$$\sum_{j=1}^n p_{ij}^\alpha = p_i^\alpha \sum_{j=1}^n p_{j,i}^\alpha, \quad i = 1, 2, \dots, m \quad (4.20)$$

$$0 \leq \alpha < 1 \implies p_{j,i}^\alpha > p_{j,i}, \quad j = 1, 2, \dots, n$$

$$\implies \sum_{j=1}^n p_{j,i}^\alpha > \sum_{j=1}^n p_{j,i} = 1 \quad (\text{because } \sum_{j=1}^n p_{j,i} = 1)$$

$$\implies \sum_{j=1}^n p_{ij}^\alpha > p_i^\alpha \quad i = 1, 2, \dots, m \text{ by (4.20)}$$

$$\implies \ln \left(\sum_{i=1}^m \sum_{j=1}^n p_{ij}^\alpha \right) > \ln \sum_{i=1}^m p_i^\alpha$$

$$\Rightarrow \frac{1}{1-\alpha} \ln \sum_{i=1}^m \sum_{j=1}^n p_{ij}^\alpha > \frac{1}{1-\alpha} \ln \sum_{i=1}^n p_i^\alpha$$

$$\Rightarrow R(P*Q) > R(P)$$

$$\text{If } \alpha > 1, p_{j,i}^\alpha < p_{j,i} \Rightarrow \ln \sum_{i=1}^m \sum_{j=1}^n p_{ij}^\alpha < \ln \sum_{i=1}^m p_i^\alpha$$

$$\Rightarrow \frac{1}{1-\alpha} \ln \sum_{i=1}^m \sum_{j=1}^n p_{ij}^\alpha > \frac{1}{1-\alpha} \ln \sum_{i=1}^m p_i^\alpha$$

$$\Rightarrow R(P*Q) > R(P)$$

so that whether $\alpha > 1$ or $0 < \alpha < 1$, $R(P*Q) > R(P)$. Similarly whether $\alpha > 1$ or $0 < \alpha < 1$, $R(P*Q) > R(Q)$ and

$$2R(P*Q) > R(P) + R(Q) \quad (4.21)$$

Generalizing for k -distributions, we obtain

$$k R(P_1 * P_2 * P_3 * \dots * P_k) > R(P_1) + R(P_2) + \dots + R(P_k)$$

so that

$$\frac{R(P_1) + R(P_2) + \dots + R(P_k)}{k} < R(P_1 * P_2 * \dots * P_k) \quad (4.22)$$

Theorem 4.3.2 : Renyi's measure of entropy is a monotonic decreasing function of α . This result has been proved by Kapur[34]. Now if we denote by $R_\alpha(P)$, the Renyi measure of entropy of order α , then

$$R_0(P) = \frac{1}{1-\alpha} \ln \sum_{i=1}^m p_i^\alpha \Big|_{\alpha=0} = \ln m \quad (4.23)$$

$$R_1(P) = \lim_{\alpha \rightarrow 1} \frac{1}{1-\alpha} \ln \sum_{i=1}^m p_i^\alpha = - \sum_{i=1}^m p_i \ln p_i \quad (4.24)$$

$$R_{\infty}(P) = \lim_{\alpha \rightarrow \infty} \frac{1}{1-\alpha} \ln \sum_{i=1}^m p_i^{\alpha} = -\ln \left(\max_i p_i \right) \quad (4.25)$$

Theorem 4.3.3 : Renyi's entropy of order zero is both sub-additive and superadditive for all probability distributions.

Proof : $R_0(P) = \ln m$, $R_0(Q) = \ln n$ and $R_0(P*Q) = \ln(mn)$ so that $R_0(P*Q) = \ln(mn) = \ln(m) + \ln(n) = R_0(P) + R_0(Q)$. Therefore the inequalities

$$R_0(P*Q) \leq R_0(P) + R_0(Q) \text{ and}$$

$$R_0(P*Q) \geq R_0(P) + R_0(Q) \text{ are both satisfied for all prob. distn.}$$

Theorem 4.3.4 : Renyi's entropy of order unity is subadditive.

Proof : By definition, Renyi's entropy of order unity is same as Shannon's measure of entropy which is known to be both sub-additive and additive.

Theorem 4.3.5 : Renyi's entropy of order ∞ is superadditive or subadditive according as

$$\max_{i,j} (p_{ij}) \lesseqgtr \max_i (p_i) \max_j (p_j) \quad (4.26)$$

Proof : $R_{\infty}(P) = -\ln(\max_i p_i)$, $R_{\infty}(Q) = -\ln(\max_j q_j)$ and

$R_{\infty}(P*Q) = -\ln(\max_{i,j} p_{ij})$ so that R_{∞} is superadditive or sub-additive according as

$$R_{\infty}(P*Q) \gtrless R_{\infty}(P) + R_{\infty}(Q)$$

i.e., according as

$$\max_{i,j} p_{ij} \lesseqgtr \max_i (p_i) \max_j (q_j).$$

We shall illustrate theorem 4.3.5 by the following example.

Example 4.3.1 : Consider the following probability distributions :

I. $p_{11} = 0.1, p_{12} = 0.5, p_{21} = 0.3, p_{22} = 0.1$

$p_1 = 0.6, p_2 = 0.4; q_1 = 0.4, q_2 = 0.6$. Then we have

$\max_{i,j}(p_{ij}) = 0.5, \max_i(p_i) = 0.6, \max_j(q_j) = 0.6$. In this case

Renyi's measure of entropy of order ∞ is subadditive.

II. $p_{11} = 0.1, p_{12} = 0.3, p_{21} = 0.3, p_{22} = 0.3$

$p_1 = 0.4, p_2 = 0.6; q_1 = 0.4, q_2 = 0.6$. Then we have

$\max_{i,j}(p_{ij}) = 0.3, \max_i(p_i) = 0.6, \max_j(q_j) = 0.6$. In this case

Renyi's entropy of order ∞ is superadditive.

III. $p_{ij} = 0.25, i = 1,2, j = 1,2; p_i = 0.5, q_j = 0.5, i=1,2, j=1,2$.

Then we have

$\max_{i,j}(p_{ij}) = 0.25, \max_i(p_i) = \max_j(q_j) = 0.5$. However in this case

Renyi's entropy of order ∞ is both subadditive and superadditive.

We shall categorize the probability distributions $P \ Q$ into two types :

Type I : for which $\max_{i,j} < \max_i(p_i) \max_j(q_j)$ and

Type II : for which $\max_{i,j} \geq \max_i(p_i) \max_j(q_j)$.

The range of values of α for which Renyi's measure of entropy of order α is subadditive or superadditive depends on whether $P \ Q$ belongs to Type I or Type II class of distributions.

For independent distributions, $p_{ij} = p_i q_j$ and as such

$$\max_{i,j} (p_{ij}) = \max_i (p_i) \max_j (q_j)$$

therefore all the product distributions are classed in Type II.

Theorem 4.3.6 : Renyi's entropy of order α is a concave function of P when $0 < \alpha < 1$ and is a pseudo-concave function of P when $\alpha > 1$.

Proof : a) $0 < \alpha < 1$. We have $\frac{d^2}{dp_i^2} (p_i^\alpha) = \alpha(\alpha-1)p_i^{\alpha-2} \Rightarrow p_i^\alpha$ is a concave function of p_i , $i = 1, 2, \dots, m$

$$\Rightarrow \sum_{i=1}^m p_i^\alpha \text{ is a concave function of } P$$

$$\Rightarrow \ln \sum_{i=1}^m p_i^\alpha \text{ is a concave function of } P.$$

$$\Rightarrow \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^m p_i^\alpha \right) \text{ is a concave function of } P$$

i.e., Renyi's entropy is a concave function of P .

When $\alpha > 1$,

$$\frac{d^2}{dp_i^2} (p_i^\alpha) = \alpha(\alpha-1) p_i^{\alpha-2} \Rightarrow p_i^\alpha \text{ is a convex function of } p_i$$

$$i = 1, 2, \dots, m$$

$$\Rightarrow \sum_{i=1}^m p_i^\alpha \text{ is a convex function of } P$$

$$\Rightarrow \ln \sum_{i=1}^m p_i^\alpha \text{ is a pseudo-convex function of } P.$$

$$\Rightarrow \frac{1}{1-\alpha} \ln \sum_{i=1}^m p_i^\alpha \text{ is a pseudo-concave function of } P.$$

i.e., Renyi's measure of entropy of order α is a pseudo-concave function of P when $\alpha > 1$.

From the above discussion it appears that

- (i) $F(\alpha) = R_\alpha(P) + R_\alpha(Q)$ and $G(\alpha) = R_\alpha(P \cdot Q)$ are both monotonic decreasing functions of α as α varies from 0 to ∞ .
- (ii) Both functions start at the common value $\ln(mn)$ at $\alpha = 0$.
- (iii) At $\alpha = 1$, $F(1) = \lim_{\alpha \rightarrow 1} F(\alpha) \geq \lim_{\alpha \rightarrow 1} G(\alpha) = G(1)$, since $R_1(P \cdot Q) \leq R_1(P) + R_1(Q)$ as Shannon's measure of entropy is subadditive. Also $F(1) = G(1)$ iff P and Q are independent, otherwise $F(\alpha) < G(\alpha)$.
- (iv) As $\alpha \rightarrow \infty$, $F(\infty) \lesssim G(\infty)$ according as $R_\infty(P \cdot Q) \lesssim R_\infty(P) + R_\infty(Q)$.
- (v) The graphs of $F(\alpha)$ and $G(\alpha)$ intersect at $\alpha = 1$ if P and Q are independent, otherwise they intersect at $\alpha = \alpha$ where $\alpha > 1$.

We have drawn the following graphs $G(\alpha) - F(\alpha)$ vs α for some experimental distributions from both the types.

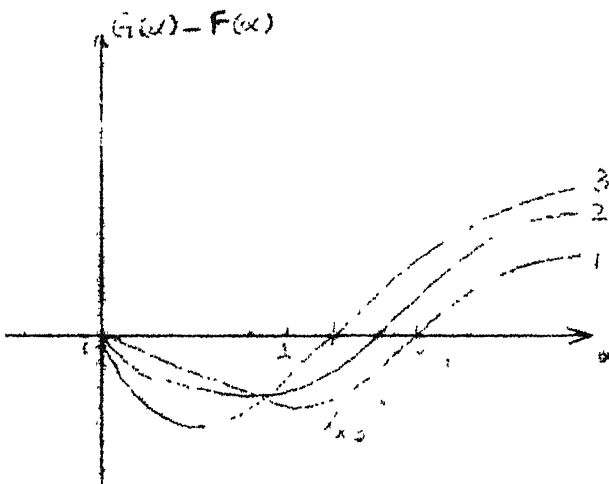


Fig. 4.3.1

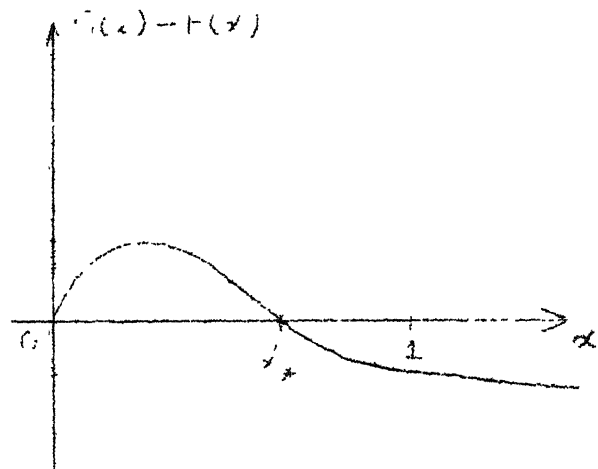


Fig. 4.3.2

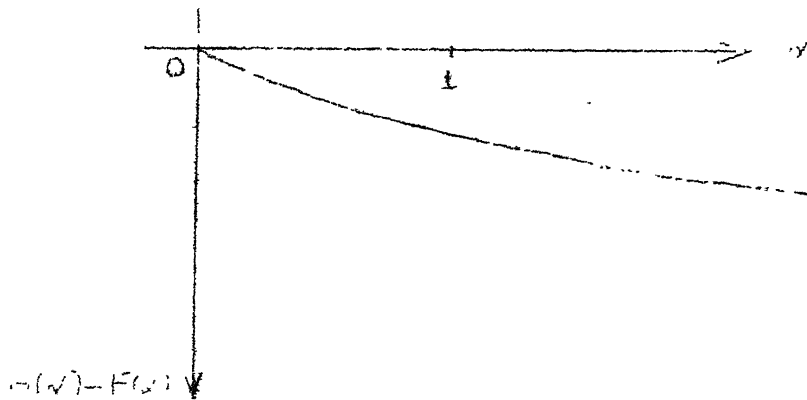


Fig. 4.3.3

Here in Fig. 4.3.1, $G(\alpha) - F(\alpha)$ intersects the α -axis at no point between 0 and 1, and intersects at α_* between 1 and ∞ for type I distributions so that for all values of α lying between 0 and α_* it is subadditive and for $\alpha > \alpha_*$ it is superadditive. In Fig. 4.3.1, we have considered the following three type I distributions :

- (1) : $P*Q = \begin{pmatrix} 0.005 & 0.095 \\ 0.285 & 0.615 \end{pmatrix}$, $P = (0.1, 0.9)$ and $Q = (0.29, 0.71)$
- (2) : $P*Q = \begin{pmatrix} 0.015 & 0.085 \\ 0.275 & 0.625 \end{pmatrix}$, $P = (0.1, 0.9)$ and $Q = (0.29, 0.71)$
- (3) : $P*Q = \begin{pmatrix} 0.025 & 0.075 \\ 0.265 & 0.635 \end{pmatrix}$, $P = (0.1, 0.9)$ and $Q = (0.29, 0.71)$

In Fig. 4.3.2 we considered a type II distribution,

$P*Q = \begin{pmatrix} 0.05 & 0.05 \\ 0.1 & 0.8 \end{pmatrix}$, $P = (0.1, 0.9)$ and $Q = (0.15, 0.85)$. Here we get an α_* lying between 0 and 1 so that for $\alpha \leq \alpha_*$, $R_\alpha(P*Q)$ is superadditive and for all $\alpha > \alpha_*$, $R_\alpha(P*Q)$ is subadditive.

Again we considered the following type II distributions :

- (1) : $P*Q = \begin{pmatrix} 0.1 & 0.5 \\ 0.3 & 0.1 \end{pmatrix}$, $P = (0.6, 0.4)$ and $Q = (0.4, 0.6)$

$$(2) : P*Q = \begin{pmatrix} 0.08 & 0.02 \\ 0.02 & 0.83 \end{pmatrix}, P = (0.1, 0.9) \text{ and } Q = (0.1, 0.9)$$

then we observe that no α_* exists in these cases and $R_\alpha(P)$ is always subadditive.

However our above discussion does not rule out the possibility of an even number of points of intersection between $G(\alpha) - F(\alpha)$ and the α -axis between 0 and 1 and an odd number of points of intersection between 1 to ∞ , for other type I distributions. However in view of the monotonic and concavity characters of $F(\alpha)$ and $G(\alpha)$ this possibility is evidently unlikely. The large number of calculations we have carried out also suggest that for Type I distributions, there is only one point of intersection α_* , where $\alpha_* > 1$.

So it has been established that for every distribution $P*Q$ there is a certain interval including $\alpha = 1$ in which $R_\alpha(P)$ is subadditive. We now tabulate some of our numerical computations.

We observe that for the first two distributions in Table 4.3.2, $R_\alpha(P)$ is subadditive for all α whereas in the case of third distribution, it is subadditive only for $\alpha \geq \alpha_*$, where $0.1 < \alpha_* < 0.2$ and superadditive for all $\alpha \leq \alpha_*$. In the case of the fourth distribution, α_* lies in the interval $(0.52, 0.55)$.

The following table comprises of the critical values of α for type I distributions :

Sl. No.	Joint distribution	Marginal distributions	Interval containing α	$G(\alpha) - F(\alpha)$
1.	0.005, 0.095	(0.1, 0.9)	(1.280, 1.285)	-0.2934×10^{-3}
	0.285, 0.615	(0.29, 0.71)		0.1735×10^{-4}
2.	0.010, 0.090	-do-	(1.225, 1.230)	-0.2215×10^{-3}
	0.280, 0.620			0.1083×10^{-4}
3.	0.015, 0.085	-do-	(1.170, 1.175)	-0.1223×10^{-3}
	0.275, 0.625			0.4011×10^{-4}
4.	0.020, 0.08	-do-	(1.115, 1.120)	-0.4835×10^{-5}
	0.270, 0.63			0.9384×10^{-4}
5.	0.025, 0.075	-do-	(1.050, 1.055)	-0.2578×10^{-4}
	0.265, 0.635			0.1597×10^{-4}
6.	0.120, 0.230	(0.35, 0.65)	(1.060, 1.065)	-0.2190×10^{-5}
	0.230, 0.420	(0.35, 0.65)		0.2563×10^{-5}
7.	0.080, 0.220	(0.3, 0.7)	(1.150, 1.155)	-0.1490×10^{-5}
	0.220, 0.480	(0.3, 0.7)		0.2644×10^{-4}

Table 4.3.1

In table 4.3.1, the value α_* lies between the two values in the 4th column. When P and Q are independent, $\alpha_* = 1$; when P and Q are nearly independent α_* is slightly greater than unity and when P and Q are far from independence α_* differs from unity significantly. In fact $(\alpha_* - 1)$ can be used as a measure of dependence between P and Q via $P*Q$.

In fact when $\alpha = 1$, $R_1(P) + R_1(Q) - R_1(P*Q)$ is itself taken as a measure of dependence. However when $\alpha > 1$,

$R_{\alpha}(P) + R_{\alpha}(Q) - R_{\alpha}(P * Q)$ can be negative and as such the value of α_* for which this vanishes for any given distribution $P * Q$ is taken as a measure of dependence.

The following table comprises of critical values of α for type II distributions :

Sl. No.	Joint Distribution	Marginal Distributions	Critical Interval	$G(\alpha) - F(\alpha)$
1.	0.1 , 0.5	(0.6 , 0.4)	Doesn't exist	$\alpha = \frac{1}{2}, -0.957 \times 10^{-1}$
	0.3 , 0.1	(0.4 , 0.6)		$\alpha = 2, -0.2862017$
				$\alpha = 0.8, -0.1470578$
				$\alpha = 0.1, -0.17759 \times 10^{-1}$
2.	0.08, 0.02	(0.1, 0.9)	Doesn't exist	$\alpha = \frac{1}{2}, -0.12406$
	0.02, 0.88	(0.1, 0.9)		$\alpha = 0.1, -0.231189 \times 10^{-1}$
				$\alpha = 0.8, -0.16858$
				$\alpha = 2.0, -0.1504919$
3.	0.1 , 0.1	(0.2 , 0.8)	(0.1, 0.2)	$\alpha = 2, -0.1174$
	0.1 , 0.7	(0.2 , 0.8)		$\alpha = \frac{1}{2}, -0.163518 \times 10^{-1}$
				$\alpha = 0.8, -0.419381 \times 10^{-1}$
				$\alpha = 0.1, +0.56605 \times 10^{-3}$
				$\alpha = 0.2, 0.96762 \times 10^{-3}$
4.	0.05, 0.05	(0.1 , 0.9)	(0.52, 0.55)	$\alpha = 0.1, 0.1081737$
	0.1 , 0.8	(0.15, 0.85)		$\alpha = 0.5, 0.21524 \times 10^{-2}$
				$\alpha = 0.52, 0.0007$
				$\alpha = 0.55, -0.00157$
				$\alpha = 0.6, -0.55936 \times 10^{-2}$
				$\alpha = 0.9, -0.31620 \times 10^{-1}$
				$\alpha = 2.0, -0.6970 \times 10^{-1}$

Table 4.3.2

4.4 Subadditivity and Superadditivity of Havrda-Charvats' measure of entropy

Theorem 4.4.1 : Havrda-Charvats' measure of entropy satisfies the inequality

$$H(P*Q) > H(P) \text{ and } H(P*Q) > H(Q). \quad (4.27)$$

Proof : It is similar to that of Theorem 4.3.1.

Theorem 4.4.2 : Havrda-Charvats' measure of entropy is a monotonic decreasing function of α .

For proof, refer Kapur [34].

Theorem 4.4.3 : Havrda-Charvats' measure of entropy of order zero is superadditive.

$$\begin{aligned} \text{Proof : } H_0(P*Q) - H_0(P) - H_0(Q) &= mn-1-(m-1)-(n-1) \\ &= (m-1)(n-1) > 0 \end{aligned}$$

because $m \geq 2, n \geq 2$.

$$\therefore H_0(P*Q) > H_0(P) + H_0(Q).$$

Theorem 4.4.4 : $H_1(P)$ is subadditive.

Proof follows immediately because $H_1(P)$ is Shannon's measure of entropy which is subadditive.

Theorem 4.4.5 : $H_\infty(P)$ is both superadditive and subadditive.

Proof : We can easily see that

$$H_\infty(P*Q) - H_\infty(P) - H_\infty(Q) = 0.$$

Theorem 4.4.6 : Havrda-Charvats' measure of entropy of order α is a concave function of P for all values of α .

Refer to Kapur [34] for proof.

Because Havrda-Charvat's measure is superadditive at $\alpha = 0$, and is subadditive when $\alpha \geq 1$, it must have changed from being superadditive to being subadditive at some α^* between 0 and 1.

The following table contains our numerical work with

$$H_{\alpha}(P*Q) - H_{\alpha}(P) - H_{\alpha}(Q) :$$

Sl. No.	Joint Distribution	Marginal Distributions	Critical interval of α	$H_{\alpha}(P*Q) - H_{\alpha}(P) - H_{\alpha}(Q)$
1.	0.005, 0.095 0.285, 0.615	(0.1, 0.9)	0.835, 0.840	0.1711×10^{-3} -0.619×10^{-3}
2.	0.010, 0.090 0.280, 0.620	(0.1, 0.9) (0.29, 0.71)	0.920, 0.925	0.641×10^{-4} -0.749×10^{-3}
3.	0.015, 0.085 0.275, 0.625	(0.1, 0.9) (0.29, 0.71)	0.960, 0.965	0.605×10^{-3} -0.251×10^{-3}
4.	0.020, 0.080 0.270, 0.630	(0.1, 0.90) (0.29, 0.71)	0.985, 0.990	0.313×10^{-3} -0.581×10^{-3}
5.	0.025, 0.075 0.265, 0.635	(0.1, 0.9) (0.29, 0.71)	0.9975, 0.9980	0.298×10^{-4} -0.745×10^{-4}
6.	0.030, 0.070 0.260, 0.640	(0.1, 0.9) (0.29, 0.71)	0.99960 0.99965	0.745×10^{-4} -0.119×10^{-3}
7.	0.150, 0.250 0.250, 0.350	(0.4, 0.6) (0.4, 0.6)	0.9980 0.9985	0.373×10^{-4} -0.198×10^{-3}
8.	0.200, 0.250 0.250, 0.300	(0.45, 0.55) (0.45, 0.55)	0.99975 0.99980	0.596×10^{-4} -0.745×10^{-4}

Table 4.4.1

From Tables 4.3.1 and 4.3.2 we deduce the following table :

Sl. No.	1	2	3	4	5	6	7
α^*	0.838	0.920	0.963	0.987	0.957	0.9576	1.0000
α_*	1.285	1.230	1.175	1.125	1.065	1.053	1.0000
$1-\alpha^*$	0.162	0.080	0.037	0.0130	0.0030	0.0024	0
α_*^{-1}	0.285	0.230	0.175	0.125	0.065	0.053	0

Table 4.4.2

We find that both $1-\alpha^*$ and α_*^{-1} decrease or increase together and either can be used as a measure of dependence, between P and Q viz $P*Q$. We describe the situation in Fig.4.4.1.

	Superadditive	Subadditive
Havrda-Charvat	α_*	1
	Subadditive	Superadditive
Renyi	1	α^*

Fig. 4.4.1.

In general the portions of subadditivity and superadditivity are given in the above figure, for type I distributions. $\alpha = 1$ is always included in the subadditivity range and $1-\alpha^*$ and α_*^{-1} can always be used as measures of dependence, α^* and α_* of course depending on $P*Q$.

For $H_\alpha(P)$ our classification of the bivariate probability distributions into type I and type II has no relevance. That is

so because irrespective of the type of the distribution $H_\alpha(P)$ is subadditive for $\alpha = 1$ and superadditive for $\alpha = 0$, thereby assuring us of atleast one α^* , $0 < \alpha^* < 1$. Though we have been unable to rule out the possibility of more than one α^* and α_* , we can reasonably consider that this possibility is highly unlikely. Even if there exist more than one α^* or α_* , we may consider α^* and α_* nearest to unity for finding the degree of dependence between P and Q in $P*Q$.

A consequence of Theorem 4.2.1(a) is that since H_α is subadditive in $(\alpha^*, 1)$ R should also be subadditive in $(\alpha^*, 1)$. Similarly we deduce from Theorem 4.2.2(a) that subadditivity of R_α in $(1, \alpha_*)$ means subadditivity of H_α there. Then in (α^*, α_*) both H_α and R_α are subadditive. Note that this holds good only in the case of type I distributions. In the case of type II distributions, if α_* exists, both α^* and α_* are less than 1. There's noway of determining which of them is larger. Thus for type I distributions, we have established a common interval for α , containing 1 where H_α and H^α are both subadditive.

4.5 Subadditivity and Superadditivity of Kapur-Aczel-Daroczy Measure of Entropy

$$H_{\alpha, \beta}(P) = \frac{1}{1-\alpha} \ln \left\{ \frac{\sum_{i=1}^n p_i^{\alpha+\beta-1}}{\sum_{i=1}^n p_i^\beta} \right\} \quad \alpha \neq 1. \quad (4.28)$$

If $\beta = 1$, (4.28) reduces to $R_\alpha(P)$. So all the results obtained in section 4.2 and 4.3 follow as a special case for $H_{\alpha, \beta}$, $\beta = 1$.

We have performed some calculations for this measure using type I distributions. Our results are tabulated below. For meaningful definition of this measure we should have either $\alpha > 1, \beta < 1$ or $\alpha < 1, \beta > 1$. Confining ourselves to this restriction we have fixed one parameter (α or β) and found the value of the other parameter up to which $H_{\alpha,\beta}$ is subadditive, for some of the type I distributions we considered for $R_\alpha(P)$ and $H_\alpha(P)$.

Distribution : $P*Q = (0.005, 0.095), (0.285, 0.615)$.

Fixed β :

β	α	$H_{\alpha,\beta}(P*Q) - H_{\alpha,\beta}(P) - H_{\alpha,\beta}(Q)$	Inferences
1.3	0.65	-0.11691×10^{-2}	For $\beta = 1.3$, $H_{\alpha,\beta}$ is subadditive for all $\alpha \leq 0.65$, superadditive for $0.70 \leq \alpha \leq 1$.
	0.70	0.40789×10^{-3}	
1.4	0.50	-0.16498×10^{-3}	For $\beta = 1.4$, $H_{\alpha,\beta}$ is subadditive for $\alpha \leq 0.5$ and superadditive for $0.55 \leq \alpha \leq 1$.
	0.55	0.14006×10^{-2}	
1.5	0.30	-0.95342×10^{-3}	For $\beta = 1.5$, $H_{\alpha,\beta}$ is subadditive for $\alpha \leq 0.3$ and superadditive for $0.35 \leq \alpha < 1$.
	0.35	0.63733×10^{-3}	
			For all $\beta > 1.5$, $H_{\alpha,\beta}$ is always superadditive for $0 \leq \alpha \leq 1$.

Table 4.5.1

Fixed α

α	β	$H_{\alpha,\beta}(P*Q) - H_{\alpha,\beta}(P) - H_{\alpha,\beta}(Q)$	Inferences
0.1	1.65	-0.18627×10^{-2}	For $\alpha = 0.1$, $H_{\alpha,\beta}$ is subadditive for $1 \leq \beta \leq 1.65$ and superadditive for $\beta \geq 1.70$.
	1.70	0.74455×10^{-3}	
0.5	1.45	-0.16486×10^{-3}	For $\alpha = 0.5$, $H_{\alpha,\beta}$ is subadditive for $1 \leq \beta \leq 1.45$ and superadditive for $\beta \geq 1.50$
	1.50	0.25758×10^{-2}	
0.9	1.20	-0.22830×10^{-2}	For $\alpha = 0.9$, $H_{\alpha,\beta}$ is subadditive for $1 \leq \beta \leq 1.2$ and superadditive for $\beta \geq 1.2$
	1.25	0.70753×10^{-3}	

Table 4.5.2

Distribution : $P*Q = (0.01, 0.09), (0.28, 0.62)$.

Fixed β :

β	α	$H_{\alpha,\beta}(P*Q) - H_{\alpha,\beta}(P) - H_{\alpha,\beta}(Q)$	Inferences
1.3	0.60	-0.83604×10^{-3}	For $\beta = 1.3$, $H_{\alpha,\beta}$ is subadditive for $0 \leq \alpha \leq 0.6$ and superadditive for $0.65 \leq \alpha \leq 1.0$.
	0.65	0.28051×10^{-3}	
1.4	0.45	-0.63866×10^{-4}	For $\beta = 1.4$, $H_{\alpha,\beta}$ is subadditive for $0 \leq \alpha \leq 0.45$ and superadditive for $0.5 \leq \alpha \leq 1$
	0.50	0.10350×10^{-2}	
1.5	0.25	-0.49169×10^{-3}	For $\beta = 1.5$, $H_{\alpha,\beta}$ is subadditive for $0 \leq \alpha \leq 0.25$ and superadditive for $0.3 \leq \alpha \leq 1.0$
	0.30	0.58626×10^{-3}	
			For $\beta > 1.5$, $H_{\alpha,\beta}$ is subadditive.

Table 4.5.3.

Fixed α :

α	β	$H_{\alpha,\beta}(P*Q) - H_{\alpha,\beta}(P) - H_{\alpha,\beta}(Q)$	Inferences
0.5	1.40	-0.099483×10^{-3}	For $\alpha = 0.5$, $H_{\alpha,\beta}$ is subadditive for $1 \leq \beta \leq 1.40$ and is superadditive for $1.45 \leq \beta$.
	1.45	0.10350×10^{-2}	
0.9	1.20	-0.54780×10^{-3}	For $\alpha = 0.9$, $H_{\alpha,\beta}$ is subadditive for $1.0 \leq \beta \leq 1.20$ and is superadditive for $1.25 \leq \beta$.
	1.25	0.15460×10^{-2}	
			For $\alpha < 0.5$ also $H_{\alpha,\beta}$ is superadditive for $\beta > 1$.

Table 4.5.4.

With that we conclude this chapter.

Chapter 5

Two Optimization Problems in Information Theory

Introduction :

A probability space is one in which a point P has n coordinates p_1, p_2, \dots, p_n where

$$p_1 \geq 0, p_2 \geq 0, \dots, p_n \geq 0 \text{ and } \sum_{i=1}^n p_i = 1 \quad (5.1)$$

so that every point represents a probability distribution. The distance of a point P from a point $Q = (q_1, \dots, q_n)$ is defined by the Kullback-Leibler 2 measure of directed divergence :

$$D(P:Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} . \quad (5.2)$$

This distance is not symmetric with respect to P and Q and it does not satisfy the triangle inequality. In spite of these two weaknesses, the geometry of the probability space characterised by the distance function (5.2) is both interesting and useful. The distance $D(P:Q)$ represents the directed divergence of P from Q and plays important role in information theory and its applications in various fields [60,32], specially through its role in Kullback's [3] principle of minimum discrimination information. $D(P:Q)$ is a convex function of both p_1, p_2, \dots, p_n and q_1, q_2, \dots, q_n and is thus very suitable for use in optimization problems.

The special functional form (5.2) is particularly useful since it involves the logarithmic function which is real only when the variable is greater than zero. This ensures that the answers we shall get to our optimization problems will be positive and will usually belong to the probability space itself.

In this chapter we consider two optimization problems. First we use (5.2) as the distance function. Later we shall use other measures of directed divergence instead of (5.2) and solve the problems. But before we introduce the problems, we consider, in the next section, some results about the geometry of the probability space.

5.1 Geometry of the Probability Space

There are n special points in the probability space. viz.

$$\begin{aligned}
 I_1 &= (1, 0, 0, \dots, 0) \\
 I_2 &= (0, 1, 0, \dots, 0) \\
 &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
 &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
 &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
 I_n &= (0, 0, \dots, 0, 1)
 \end{aligned}
 \tag{5.3}$$

Corresponding to n degenerate distributions. Each of these distributions represents a state of certainty. Also

$$D(I_k, P) = -\ln p_k, \quad k = 1, 2, \dots, n \tag{5.4}$$

represents the distance of the k^{th} state of certainty from the probability distribution P . It depends only on p_k and in some

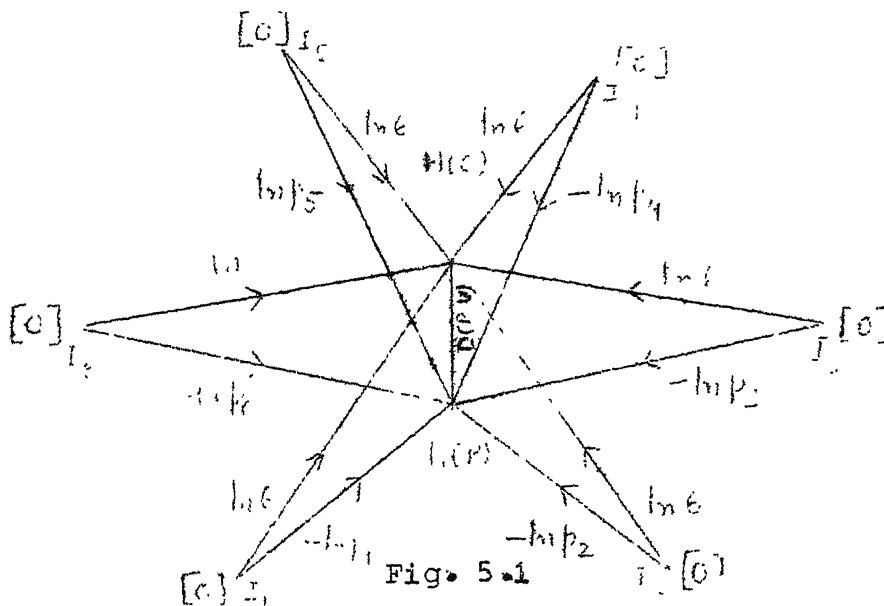
sense it gives the measure of uncertainty associated with the k^{th} outcome of the distribution P . The average of all such uncertainties is

$$H(P) = - \sum_{i=1}^n p_i \ln p_i \quad (5.5)$$

which gives a measure of uncertainty associated with the prob. distn. P . This is defined as entropy of P .

Uncertainty is minimum when P coincides with any of the points (5.3) and is maximum when $p_1 = p_2 = \dots = p_n = \frac{1}{n}$ and the maximum value is equal to $\ln n$.

Fig. 5.1 gives the geometry of the probability space for $n = 6$ where the numbers within squares give entropies and numbers along the directed lines give directed divergences



The minimum distance between two points P and Q is zero and arises when $P = Q$. For finding the maximum distance we keep

Q fixed and note that $\sum_{i=1}^n p_i \ln p_i/q_i$ is a convex function of p_1, \dots, p_n whose maximum value can occur at one of the n points I_1, I_2, \dots, I_n . Its values at these points are $-\ln q_1, \dots, -\ln q_n$ so that the maximum value of $D(P:Q)$ is $-\ln q_{\min}$ where q_{\min} is the minimum of q_1, q_2, \dots, q_n . As the $q_{\min} \rightarrow 0$, this distance approaches infinity and as such the distance between two points in our geometry can be arbitrarily large.

If P and Q are two distributions then the set of distns. $\lambda Q + (1-\lambda)P$, $0 \leq \lambda \leq 1$, will be said to constitute a line segment. If $\lambda > 1$ or $\lambda < 0$ this may not be a prob. distn. and as such this does not give a straight line. This is unlike the Euclidean case.

The set of prob. distns. P satisfying

$$\sum_{i=1}^n p_i \ln \frac{p_i}{q_i} = K \quad (5.6)$$

where $Q = (q_1, \dots, q_n)$ is a fixed prob. distn. with each $q_i > 0$ gives a 'hypersphere' in the probability space, with 'centre' Q and 'radius' K . However this defines a hypersphere only when $K < -\ln q_{\min}$. If $K = -\ln q_{\min}$ there is a single point and if $K > -\ln q_{\min}$ the hypersphere is imaginary.

Similarly

$$\sum_{i=1}^n p_i \ln \frac{p_i}{q_i} + \sum_{i=1}^n p_i \ln \frac{p_i}{r_i} = K \quad (5.7)$$

where $Q = (q_1, \dots, q_n)$ with $q_i > 0 \forall i = 1, \dots, n$ and $R = (r_1, \dots, r_n)$, $r_i > 0 \forall i = 1, \dots, n$ are any fixed distributions,

defines a hyper-ellipsoid with foci at Q and R. This will not exist for all values of K. We investigate in section 4 the values of K for which prob. distn. satisfying (5.7) exist.

5.2 First Optimization Problem

This problem is concerned with finding the maximum and minimum values of

$$\theta = D(P:R) - D(P:Q) \quad (5.8)$$

where Q and R are fixed distributions for which each component is greater than zero.

The motivation to this problem arises from a problem solved by Kullback [3] who found the distn. P which minimized $D(P:R)$ s.to. $D(P:R) - D(P:Q) = \theta$ where θ is a fixed constant. He discussed the solution for θ lying between 0 and 1, but his solution is valid for a larger range of values of θ . We want to find this range precisely

$$D(P:R) - D(P:Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{r_i} - \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} = \sum_{i=1}^n p_i \ln \frac{q_i}{r_i} \quad (5.9)$$

(5.9) is to be maximized s.to $\sum_{i=1}^n p_i = 1$. Its maximum value is

$$\max (\ln \frac{q_1}{r_1}, \ln \frac{q_2}{r_2}, \dots, \ln \frac{q_n}{r_n}). \quad (5.10)$$

Thus

$$(\theta)_{\max} = D(P:R) - D(P:Q)_{\max} = \ln \left(\frac{q_i}{r_i} \right)_{\max} \quad (5.11)$$

Since $\sum_{i=1}^n q_i = 1$, $\sum_{i=1}^n r_i = 1$ and $Q \neq R$ there are one or more q_i 's which are greater than the corresponding r_i 's and so $(q_i/r_i)_{\max} > 1$ and $(\theta)_{\max} > 0$. Moreover,

$$\ln \left(\frac{q_i}{r_i} \right)_{\max} = \sum_{j=1}^n q_j \ln \left(\frac{q_i}{r_i} \right)_{\max} \geq \sum_{j=1}^n q_j \ln \frac{q_j}{r_j} = D(Q:R) \quad (5.12)$$

Thus we have

$$[D(P:R) - D(P:Q)]_{\max} \geq D(Q:R). \quad (5.13)$$

(5.13) is a weaker form of triangle inequality according to which in Euclidean geometry

$$D(P:R) - D(P:Q) \geq D(Q:R). \quad (5.14)$$

Now we minimize $D(P:R) - D(P:Q)$ subject to $\sum_{i=1}^n p_i = 1$. We write

$$D(P:R) - D(P:Q) = - \sum_{i=1}^n p_i \ln \frac{r_i}{q_i} \quad (5.15)$$

$$[D(P:R) - D(P:Q)]_{\min} = -\ln \left(\frac{r_i}{q_i} \right)_{\max} \quad (5.16)$$

$$\text{Again } -\ln \left(\frac{r_i}{q_i} \right)_{\max} = - \sum_{j=1}^n r_j \left(\ln \frac{r_i}{q_i} \right)_{\max} \leq - \sum_{j=1}^n r_j \left(\ln \frac{r_j}{q_j} \right) = -D(R:Q) \quad (5.17)$$

$$\text{Thus } D(P:R) - D(P:Q)_{\min} \leq -D(R:Q). \quad (5.18)$$

In fact inequality (5.18) can be deduced from (5.13).

For, by interchanging Q and R in (5.13), we get

$$[D(P:Q) - D(P:R)]_{\max} \geq D(R:Q)$$

$$\text{or } -[D(P:Q) - D(P:R)]_{\max} \leq -D(R:Q)$$

$$\text{or } [D(P:R) - D(P:Q)]_{\min} \leq -D(R:Q)$$

which is same as (5.18).

Thus the max. value of θ is $\ln \left(\frac{q_i}{r_i} \right)_{\max}$ which is $> D(Q:R)$ and the minimum value is $\ln(q_i/r_i)_{\min}$ which is $< -D(R:Q)$. The maximum value arises for that degenerate distribution P for which that component is 1 for which $\left(\frac{q_i}{r_i} \right)$ is maximum and the minimum value arises for that degenerate distribution for which that component is unity for which (q_i/r_i) minimum.

Thus Kullback's problem can be solved when θ lies between $-\ln \left(\frac{q_i}{r_i} \right)_{\min}$ and $\ln (q_i/r_i)_{\max}$.

An alternative proof of the above result can be obtained by attempting to solve Kullback's problem. We do it in the next section.

5.3 An alternative proof of the results regarding the Maximum and Minimum Values of θ

We minimize

$$\sum_{i=1}^n p_i \ln \frac{p_i}{r_i} \quad (5.20)$$

subject to

$$\sum_{i=1}^n p_i \ln \frac{p_i}{r_i} - \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} = \theta \quad (5.21)$$

and

$$\sum_{i=1}^n p_i = 1 \quad \text{to get}$$

$$p_i = \frac{q_i^\alpha r_i^{1-\alpha}}{\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha}} \quad (5.22)$$

where α is determined from the equation

$$\frac{\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \ln \frac{q_i}{r_i}}{\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha}} = \theta \quad (5.23)$$

$$\text{Let } f(\alpha) = \sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \quad (5.24)$$

$$\text{Then we have } f'(\alpha) = \sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \ln \frac{q_i}{r_i} \quad (5.25)$$

$$f''(\alpha) = \sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \left(\ln \frac{q_i}{r_i} \right)^2 \quad (5.26)$$

$$f(0) = 1, \quad f(1) = 1 \quad (5.27)$$

$$f'(0) = \sum_{i=1}^n r_i \ln \frac{q_i}{r_i} = -D(R:Q) \leq 0 \quad \because D(R:Q) \geq 0 \quad (5.28)$$

$$f'(1) = \sum_{i=1}^n q_i \ln \frac{q_i}{r_i} = D(R:Q) \geq 0 \quad (5.29)$$

and $f(\alpha)$ is a convex function with graph given in Fig. 5.2 and α_0 is determined from

$$f'(\alpha)/f(\alpha) = \theta \quad (5.30)$$

Also $f(\alpha)$ has a minimum at $\alpha = \alpha_0$ where $0 < \alpha_0 < 1$.

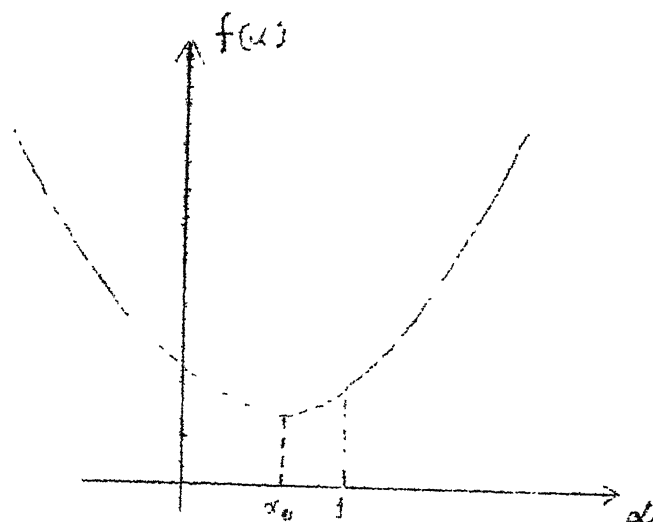


Fig. 5.2

$$\begin{aligned}
 \text{Now } \frac{d\theta}{d\alpha} &= \frac{f(\alpha) f''(\alpha) - f'^2(\alpha)}{f^2(\alpha)} \\
 &= \frac{\left(\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \right) \left(\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \left(\ln \frac{q_i}{r_i} \right)^2 \right) - \left(\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \ln \frac{q_i}{r_i} \right)^2}{\left(\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \right)^2} \geq 0
 \end{aligned} \tag{5.31}$$

Here we have used Cauchy's inequality. Therefore θ increases with α and

$$\theta_{\max} = \lim_{\alpha \rightarrow \infty} \frac{f'(\alpha)}{f(\alpha)}, \quad \theta_{\min} = \lim_{\alpha \rightarrow -\infty} \frac{f'(\alpha)}{f(\alpha)} \tag{5.32}$$

so that

$$\theta_{\max} = \lim_{\alpha \rightarrow \infty} \frac{\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \ln \frac{q_i}{r_i}}{\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha}} = \ln \left(\frac{q_i}{r_i} \right)_{\max} \tag{5.33}$$

and

$$\theta_{\min} = \lim_{\alpha \rightarrow -\infty} \frac{\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha} \ln \frac{q_i}{r_i}}{\sum_{i=1}^n q_i^\alpha r_i^{1-\alpha}} = -\ln \left(\frac{r_i}{q_i} \right)_{\max} \tag{5.34}$$

$$\text{Also } (\theta)_{\alpha=1} = \frac{f'(1)}{f(1)} = D(Q;R) \quad (5.35)$$

$$(\theta)_{\alpha=0} = \frac{f'(0)}{f(1)} = -D(R:Q) \quad (5.36)$$

Since θ is a monotonic increasing function of α , $\lim_{\alpha \rightarrow \infty} \theta > f(1)$ and $\lim_{\alpha \rightarrow -\infty} \theta < f(0)$ and so all the results of the last section follow.

5.4 Second Optimization Problem

This problem is concerned with finding the maximum and minimum values of

$$\varphi = \lambda_1 D(P:Q_1) + \lambda_2 D(P:Q_2) + \dots + \lambda_m D(P:Q_m) \quad (5.37)$$

$$\text{where } \lambda_j > 0 \quad \forall j = 1, \dots, m, \quad \sum_{j=1}^m \lambda_j = 1 \quad (5.38)$$

$$\text{and } Q_j = (q_{j1}, \dots, q_{jn}), \quad j = 1, \dots, m \quad (5.39)$$

are given probability distributions with each $q_{ji} > 0$. Thus this problem is concerned with finding the maximum and minimum values of the weighted sum of directed divergences of P from Q_1, Q_2, \dots, Q_m .

$$\begin{aligned} \text{Now } &= \sum_{j=1}^m \lambda_j \sum_{i=1}^n p_i \ln \frac{p_i}{q_{ji}} = \sum_{j=1}^m \lambda_j \sum_{i=1}^n p_i \ln p_i \\ &\quad - \sum_{j=1}^m \lambda_j \sum_{i=1}^n p_i \ln q_{ji} \\ &= \sum_{i=1}^n p_i \ln \frac{p_i}{\bar{q}_i} = \sum_{i=1}^n p_i \ln \left(\frac{p_i}{\bar{q}_i / \beta} \right) - \ln \beta \quad (5.40) \end{aligned}$$

$$\text{where } \bar{q}_i = \prod_{j=1}^m q_{ji} \quad (5.41)$$

is the weighted geometric mean of the i^{th} components of the prob. distns. Q_1, Q_2, \dots, Q_m and

$$\beta = \sum_{i=1}^n \bar{q}_i \quad (5.42)$$

so that $\frac{\bar{q}_1}{\beta}, \frac{\bar{q}_2}{\beta}, \dots, \frac{\bar{q}_n}{\beta}$ is a probability distn. The minimum and maximum values of $\sum_{i=1}^n p_i \ln \left(\frac{p_i}{\bar{q}_i/\beta} \right)$ are zero and $\ln \frac{\beta}{(\bar{q}_i)_{\min}}$.

Thus the minimum value of ϕ is $-\ln \beta$ and it occurs when $p_i = \bar{q}_i / \sum_{i=1}^n \bar{q}_i$ and the maximum value of ϕ is $\ln \frac{1}{(\bar{q}_i)_{\min}}$ and it occurs when P is the degenerate distn. whose only non-zero component is that for which \bar{q}_i is minimum. Now

$$\beta = \sum_{i=1}^n \bar{q}_i = \sum_{i=1}^n \prod_{j=1}^m q_{ji} < \sum_{j=1}^m \left(\sum_{i=1}^n g_{ji} \right)^{\lambda_j} = 1, -\ln \beta > 0 \quad [GM < AM] \quad (5.43)$$

so that as expected, both the maximum and the minimum values of ϕ are positive.

Now we shall use other measures of directed divergence as distance functions in solving these two optimization problems.

5.5 First Problem Using Generalized Measures of Directed Divergence :

5.5.1 Havrda-Charvat Measure of Directed Divergence

$$D^\alpha(P:Q) = \frac{1}{\alpha-1} \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} - 1 \quad \alpha \neq 1 \quad (5.44)$$

is the new distance measure in the probability space. We shall maximize and minimize $\theta_0 = \frac{1}{\alpha-1} \sum_{i=1}^n p_i^\alpha (q_i^{1-\alpha} - q_i^{1-\alpha})$ (5.45)

subject to $\sum_{i=1}^n p_i = 1$.

Consider $(r_i^{1-\alpha} - q_i^{1-\alpha})$, $i = 1, 2, \dots, n$. Let $i = k$ be the index for which $(r_i^{1-\alpha} - q_i^{1-\alpha})$ attains the maximum value. Viz.

$$(r_k^{1-\alpha} - q_k^{1-\alpha}) = \max_i (r_i^{1-\alpha} - q_i^{1-\alpha}) \quad (5.46)$$

Then we have

$$(r_k^{1-\alpha} - q_k^{1-\alpha}) \geq (r_i^{1-\alpha} - q_i^{1-\alpha}) \quad \forall i = 1, \dots, n$$

$$\Rightarrow \sum_{i=1}^n p_i^\alpha (r_k^{1-\alpha} - q_k^{1-\alpha}) \geq \sum_{i=1}^n p_i^\alpha (r_i^{1-\alpha} - q_i^{1-\alpha}) \quad (5.47)$$

Case (i) : Now let $\alpha \geq 1$, then from (5.47) we get

$$\frac{1}{(\alpha-1)} \sum_{i=1}^n p_i^\alpha (r_k^{1-\alpha} - q_k^{1-\alpha}) \geq \frac{1}{(\alpha-1)} \sum_{i=1}^n p_i^\alpha (r_i^{1-\alpha} - q_i^{1-\alpha}). \quad (5.48)$$

$$\text{But for } \alpha \geq 1, \quad p_i^\alpha \leq 1. \therefore \frac{r_k^{1-\alpha} - q_k^{1-\alpha}}{(\alpha-1)} \geq \frac{(r_k^{1-\alpha} - q_k^{1-\alpha})}{(\alpha-1)} \sum_{i=1}^n p_i^\alpha. \quad (5.49)$$

Now from (5.48) and (5.49) we get for $\alpha \geq 1$

$$\max(\theta_0) = \frac{r_k^{1-\alpha} - q_k^{1-\alpha}}{(\alpha-1)} \text{ and it occurs for } P = I_k.$$

Case (ii) : Now let $0 \leq \alpha \leq 1$, and let $i = s$ be the value of the index for which $(r_i^{1-\alpha} - q_i^{1-\alpha})$ is minimum. That is we have

$$(r_s^{1-\alpha} - q_s^{1-\alpha}) \leq (r_i^{1-\alpha} - q_i^{1-\alpha}) \quad \forall i = 1, \dots, n$$

$$\text{or } \sum_{i=1}^n p_i^\alpha (r_s^{1-\alpha} - q_s^{1-\alpha}) \leq \sum_{i=1}^n p_i^\alpha (r_i^{1-\alpha} - q_i^{1-\alpha}) \quad (5.50)$$

Now because $0 \leq \alpha < 1$, we have $(\alpha-1) < 0$ and $\sum_{i=1}^n p_i^\alpha \geq 1$. From (5.50) we get

$$\frac{1}{\alpha-1} \sum_{i=1}^n p_i^\alpha (r_s^{1-\alpha} - q_s^{1-\alpha}) \geq \frac{1}{\alpha-1} \sum_{i=1}^n p_i^\alpha (r_i^{1-\alpha} - q_i^{1-\alpha}) \quad \text{and} \quad (5.51)$$

$$\frac{r_s^{1-\alpha} - q_s^{1-\alpha}}{\alpha-1} \geq \sum_{i=1}^n p_i^\alpha \frac{(r_i^{1-\alpha} - q_i^{1-\alpha})}{(\alpha-1)} \quad (5.52)$$

From (5.51) and (5.52) we get for $0 \leq \alpha < 1$ ($\theta_o \max$) =

$$\frac{r_s^{1-\alpha} - q_s^{1-\alpha}}{\alpha-1} \quad \text{and it occurs for } P = I_s.$$

Now we shall find the minimum value of θ_o . Consider the following two cases :

$$\text{Case (iii) : Let } \alpha > 1 \text{ and } \min_i \left(\frac{r_i^{1-\alpha} - q_i^{1-\alpha}}{\alpha-1} \right) = \left(\frac{r_m^{1-\alpha} - q_m^{1-\alpha}}{\alpha-1} \right) \leq 0. \quad (5.53)$$

Here we note that $\sum_{i=1}^n r_i = \sum_{i=1}^n q_i = 1$ and therefore some q_i 's are greater than the corresponding r_i 's and the rest of q_i 's are less than the corresponding r_i 's. Therefore $\left(\frac{r_m^{1-\alpha} - q_m^{1-\alpha}}{\alpha-1} \right)$ is non-positive. Therefore we have from (5.53)

$$\sum_{i=1}^n p_i^\alpha \frac{r_i^{1-\alpha} - q_i^{1-\alpha}}{(\alpha-1)} \geq \left(\sum_{i=1}^n p_i^\alpha \right) \left(\frac{r_m^{1-\alpha} - q_m^{1-\alpha}}{(\alpha-1)} \right) \geq \frac{r_m^{1-\alpha} - q_m^{1-\alpha}}{(\alpha-1)}. \quad (5.54)$$

The minimum value of θ_0 is $(\frac{r_m^{1-\alpha} - q_m^{1-\alpha}}{\alpha-1})$ and it occurs $P = I_m$.

Case (iv) : Let $0 \leq \alpha < 1$ and $\max_i (\frac{r_i^{1-\alpha} - q_i^{1-\alpha}}{\alpha-1}) = \frac{r_M^{1-\alpha} - q_M^{1-\alpha}}{(\alpha-1)} \geq 0$.

(5.55)

Therefore

$$\sum_{i=1}^n p_i^\alpha (\frac{r_i^{1-\alpha} - q_i^{1-\alpha}}{\alpha-1}) \geq (\sum_{i=1}^n p_i^\alpha) (\frac{r_M^{1-\alpha} - q_M^{1-\alpha}}{\alpha-1}) \geq n^{1-\alpha} \frac{r_M^{1-\alpha} - q_M^{1-\alpha}}{(\alpha-1)} \quad (5.56)$$

because for $0 \leq \alpha < 1$, $\sum_{i=1}^n p_i^\alpha$ is convex and hence $P = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ minimizes $\sum_{i=1}^n p_i^\alpha$. But this value is never attained by θ_0 . So we have been able to obtain only a lower bound for the minimum of θ_0 when $0 \leq \alpha \leq 1$. We shall now summarise our results of subsection 5.5.1.

$$\text{Max}(\theta_0) : (\alpha > 1) = \frac{1}{\alpha-1} \max_i (r_i^{1-\alpha} - q_i^{1-\alpha})$$

$$(0 \leq \alpha \leq 1) = \frac{1}{\alpha-1} \min_i (r_i^{1-\alpha} - q_i^{1-\alpha})$$

$$\text{Min}(\theta_0) : (\alpha > 1) = \min_i (\frac{r_i^{1-\alpha} - q_i^{1-\alpha}}{\alpha-1})$$

$$(0 \leq \alpha < 1) \geq n^{1-\alpha} \max_i (\frac{r_i^{1-\alpha} - q_i^{1-\alpha}}{\alpha-1}).$$

5.5.2 Kapur's Measure of Directed Divergence :

$$\text{Here } D(P:Q) = D_K(P:Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} - \frac{1}{a} \sum_{i=1}^n (1+ap_i) \ln \frac{(1+ap_i)}{(1+aq_i)} \quad (5.57)$$

$$\text{and we have } \theta_K = \sum_{i=1}^n p_i \{ \ln \frac{q_i}{r_i} - \ln \frac{1+aq_i}{1+ar_i} \} - \frac{1}{a} \sum_{i=1}^n \ln \frac{1+aq_i}{1+ar_i} \quad (5.58)$$

For our purposes, the second term in (5.58) is a constant. Therefore we rewrite

$$\theta_k = \sum_{i=1}^n p_i \left\{ \ln \frac{q_i}{r_i} - \ln \frac{1+aq_i}{1+ar_i} \right\} - A \quad (5.59)$$

We can straight away conclude that

$$\max(\theta_k) = \max_i \left\{ \ln \frac{q_i(1+ar_i)}{r_i(1+aq_i)} \right\} - A \quad (5.60)$$

and

$$\min(\theta_k) = \min_i \left\{ \ln \frac{q_i(1+ar_i)}{r_i(1+aq_i)} \right\} - A.$$

We know that as $a \rightarrow 0$ (5.57) approaches Kullback-Leibler measure of directed divergence, (5.2). Now we shall show that as $a \rightarrow 0$, our results here approach our results obtained in section 5.2.

$$\begin{aligned} \lim_{a \rightarrow 0} A &= \lim_{a \rightarrow 0} \frac{1}{a} \sum_{i=1}^n \ln \left(\frac{1+ar_i}{1+aq_i} \right) \\ &= \lim_{a \rightarrow 0} \sum_{i=1}^n \left(\frac{1+aq_i}{1+ar_i} \right) \frac{(1+ar_i)q_i - (1+aq_i)r_i}{(1+aq_i)^2} \\ &= \sum_{i=1}^n (q_i - r_i) = 0. \end{aligned}$$

$\lim_{a \rightarrow 0} \max(\theta_k) = \max_i \ln \left(\frac{q_i}{r_i} \right)$ and $\lim_{a \rightarrow 0} \min(\theta_k) = \min_i \ln \left(\frac{q_i}{r_i} \right)$ which are same as our results obtained in 5.2.

5.6 Second Problem using Generalized Measures of Directed Divergence

5.6.1 Havrda-Charvats' Measure of Directed Divergence :

Here we are concerned with finding the maximum and

minimum of

$$\varphi_0 = \sum_{j=1}^m \lambda_j D^\alpha(P:Q_j)$$

where $Q_j = (q_{j1}, \dots, q_{jn})$, $j = 1, 2, \dots, m$. We have

$$\begin{aligned} &= \frac{1}{\alpha-1} \sum_{j=1}^m \lambda_j \left(\sum_{i=1}^n p_i^\alpha q_{ji}^{1-\alpha-1} \right) \\ &= \frac{1}{\alpha-1} \sum_{i=1}^n p_i^\alpha \left(\sum_{j=1}^m \lambda_j q_{ji}^{1-\alpha-1} \right) \\ &= \frac{\left(\sum_{i=1}^n \bar{q}_i \right)^{1-\alpha}}{\alpha-1} \left(\sum_{i=1}^n p_i^\alpha \left(\frac{\bar{q}_i}{\sum_{i=1}^n \bar{q}_i} \right)^{1-\alpha} - 1 \right) \end{aligned} \quad (5.61)$$

where $\bar{q}_i = \left(\sum_{j=1}^m \lambda_j q_{ji}^{1-\alpha} \right)^{1/(1-\alpha)}$. From (5.61) we find that is a constant times multiplied Havrda-Charvats' directed divergence of P from $\bar{Q} = (\bar{q}_1, \bar{q}_2, \dots, \bar{q}_n)$ where $\bar{q}_1 = (\bar{q}_1 / \sum_{i=1}^n \bar{q}_i)$.

Therefore we can easily see that the minimum value of is zero and the maximum is obtained as follows :

Case (i) : Let $\alpha > 1$. Then we have, because $x \rightarrow x^{1-\alpha}$ is a decreasing function, if $\bar{q}_k = \min_i \bar{q}_i$

$$(\bar{q}_k)^{1-\alpha} \geq \bar{q}_i^{1-\alpha} \quad \forall i = 1, 2, \dots, n \quad (5.62)$$

$$\Rightarrow \left(\sum_{i=1}^n p_i^\alpha \right) \bar{q}_k^{1-\alpha} \geq \sum_{i=1}^n p_i^\alpha \bar{q}_i^{1-\alpha}$$

$$\Rightarrow (\bar{q}_k^{1-\alpha}) \geq \left(\sum_{i=1}^n p_i^\alpha \right) (\bar{q}_k^{1-\alpha}) \geq \sum_{i=1}^n p_i^\alpha \bar{q}_i^{1-\alpha} \quad \text{because } \left(\sum_{i=1}^n p_i^\alpha \right) \leq 1$$

$$\Rightarrow (\bar{q}_k)^{1-\alpha} - 1 \geq \sum_{i=1}^n p_i^\alpha \bar{q}_i^{1-\alpha} - 1. \quad (5.63)$$

Now because $(\alpha-1)$ is greater than zero we have from (5.63) that

$$\frac{(\bar{q}_k^{1-\alpha} - 1)}{\alpha-1} \geq \frac{\sum_{i=1}^n p_i^\alpha \bar{q}_i^{1-\alpha} - 1}{(\alpha-1)} . \quad (5.64)$$

There from (5.61) and (5.64) we can easily observe that the max. value of φ_0 is given by

$$\left(\sum_{i=1}^n \bar{q}_i \right)^{1-\alpha} \frac{(\bar{q}_k^{1-\alpha} - 1)}{\alpha-1} \quad (5.65)$$

and it occurs for $P = I_k$.

Case (ii) : $0 < \alpha < 1$. Here we have $(1-\alpha) > 0$ and therefore $x \rightarrow x^{1-\alpha}$ is an increasing function. Therefore we have, instead of (5.63)

$$(\bar{q}_k)^{1-\alpha} - 1 \leq \sum_{i=1}^n p_i^\alpha (\bar{q}_k)^{1-\alpha} - 1 . \quad (5.66)$$

But now $(\alpha-1) < 0$. Therefore from (5.66) we get

$$\frac{(\bar{q}_k)^{1-\alpha} - 1}{\alpha-1} \geq \frac{\sum_{i=1}^n p_i^\alpha (\bar{q}_k)^{1-\alpha} - 1}{(\alpha-1)} . \quad (5.67)$$

Now from (5.61) and (5.67) we again get the max. of φ_0 as given in (5.65).

$$\text{Therefore for all } \alpha, \max_P = \frac{(\sum \bar{q}_i)^{1-\alpha}}{(\alpha-1)} [(\bar{q}_k)^{1-\alpha} - 1] .$$

We can easily see that our results of this subsection coincide with those in section 5.3 when α is made to approach 1.

5.6.2 Ferrari's Measure of Directed Divergence

$$D_\mu (P:Q) = \frac{1}{\mu} \sum_{i=1}^n (1 + \mu p_i) \ln \left(\frac{1 + \mu p_i}{1 + \mu q_i} \right) \quad \mu > 0$$

$$\begin{aligned}
 \text{Here } \varphi_{\mu} &= \frac{1}{\mu} \sum_{j=1}^m \lambda_j \sum_{i=1}^n (1 + \mu p_i) \ln \left(\frac{1 + \mu p_i}{1 + \mu q_{ji}} \right) \\
 &= \frac{1}{\mu} \sum_{i=1}^n (1 + \mu p_i) \ln(1 + \mu p_i) - \frac{1}{\mu} \sum_{i=1}^n (1 + \mu p_i) \ln \left(\prod_{j=1}^m (1 + \mu q_{ji})^{\lambda_j} \right) \\
 &\quad (5.68)
 \end{aligned}$$

Now we denote by $\bar{q}_i = \frac{\prod_{j=1}^m (1 + \mu q_{ji})^{\lambda_j}}{\sum_{i=1}^n \prod_{j=1}^m (1 + \mu q_{ji})^{\lambda_j}}$ then $\{\bar{q}_i\}$ denotes a

probability distribution. Then from (5.58) we get

$$\begin{aligned}
 &= \left(\frac{n+\mu}{\mu} \right) \sum_{i=1}^n \left(\frac{1+\mu p_i}{n+\mu} \right) \ln \frac{(1+\mu p_i)/(n+\mu)}{\bar{q}_i} - \left(\frac{n+\mu}{\mu} \right) \ln \left[\frac{n+\mu}{\sum_{i=1}^n \prod_{j=1}^m (1+\mu q_{ji})^{\lambda_j}} \right] . \\
 &\quad (5.69)
 \end{aligned}$$

Now we see from (5.69) that the second term of φ_{μ} is a constant and the first term can be represented as $\left(\sum_{i=1}^n \bar{p}_i \ln \frac{\bar{p}_i}{\bar{q}_i} \right) \left(\frac{n+\mu}{\mu} \right)$. The minimum and maximum values of the first term of (5.69) are zero and $\ln \left(\frac{1}{\bar{q}_{\min i}} \right)$ respectively. From this we deduce the minimum and

maximum values of φ_{μ} to be :

$$- \left(\frac{n+\mu}{\mu} \right) \ln \left(\frac{n+\mu}{\sum_{i=1}^n \prod_{j=1}^m (1 + \mu q_{ji})^{\lambda_j}} \right) \text{ and }$$

$$\left(\frac{n+\mu}{\mu} \right) \ln \left(\frac{1}{\bar{q}_{\min i}} \right) - \left(\frac{n+\mu}{\mu} \right) \ln \left(\frac{n+\mu}{\sum_{i=1}^n \prod_{j=1}^m (1 + \mu q_{ji})^{\lambda_j}} \right)$$

respectively.

5.7 Geometric Interpretation

In Euclidean space the locus of a point P which moves so that $PA - PB = C$ where A and B are fixed points and C is a constant is a hyperbola provided $C < AB$. In the case of similar problem in probability space, θ can be greater than $D(Q:R)$.

Again in Euclidean space the locus of a point, the sum of weighted distances from which to m given points is constant is an m -ellipse which is a closed convex hyper-surface and this exists only if the constant is greater than a certain critical value, but there is no upper limit to the value of this constant. On the other hand in the corresponding problem in probability space, the corresponding constant has both upper and lower bounds. The Steiner problem in Euclidean space has in general no easy solution, while the corresponding Steiner problem in probability space has an elegant solution.

One operational significance of the second problem can be as follows : m individuals whose relative importances are given by $\lambda_1, \lambda_2, \dots, \lambda_m$ have given their assessed proportions of allotments to different aspects of some work in the form of Q_1, Q_2, \dots, Q_m , then we can find an allotment which is closest to these allotments.

With that we conclude this chapter. In the next chapter we shall consider some more optimization problems in information theory.

Chapter 6

Some more Optimization Problems

Introduction. Let $Q = (q_1, q_2, \dots, q_n)$ and $R = (r_1, r_2, \dots, r_n)$ be any two given probability distributions with $q_i > 0$, $r_i > 0 \forall i=1, 2, \dots, n$ and $\sum_{i=1}^n q_i = \sum_{i=1}^n r_i = 1$.

Let $D^*(Q:R)$ be the Kullback-Leiblers' measure of directed divergence. Then Kapur [35] considered the following optimization problems:

- (1) Out of all probability distributions which are equally distant from Q and R find that distribution which is nearest to Q (or R).
- (2) Out of all distributions which are at a distance a from R , find that distribution which is closest to Q .
- (3) Out of all distributions whose distances from Q and R are in the ratio $1:k$ find that distribution which is closest to Q .
- (4) Out of all distributions whose sum of distances from Q and R is b , find that distribution which is closest to Q .
- (5) Find the distribution P from which the sum of directed divergences to k given distributions Q_1, Q_2, \dots, Q_k is minimum.

- (6) Out of all distributions from which the sum of weighted directed divergences to Q_1, Q_2, \dots, Q_k is constant, find the distribution closest to Q_0 .
- (7) Out of all distributions which are equally close to Q and R , find the distribution which is closest to another distribution S .

Here the distance from P to Q means, as in the previous chapter, the Kullback-Leibler directed divergence of P from Q .

Kapur [35] has solved the seven optimization problems and has also given solutions to their equivalents with Euclidean distances, which are well known. Here in this chapter we deal with these problems using Havrda-Charvats' measure

$$D^{\alpha}(P:Q) = \frac{1}{\alpha-1} \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} - 1$$

as the distance from P to Q .

We obtain the solutions to these problems by making use of Lagrange's method of multipliers for minimizing the distances. We consider the special cases of $\alpha = \frac{1}{2}$ and $\alpha = \frac{2}{3}$ to simplify the expressions. In some cases we present numerical solutions to these problems with having the given distributions assigned some numerical values.

6.1 Problem 1. Find a probability distribution P which is nearest to the distribution Q out of all distributions which are equidistant from the distributions Q and R .

Minimize:

$$D_{\alpha}(P:Q) = \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i^{\alpha} q_i^{1-\alpha} - 1 \right\} \text{ subject to } \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i^{\alpha} (q_i^{1-\alpha} - r_i^{1-\alpha}) = 0$$

By applying Lagrange's multipliers method we get

$$p_i = \frac{\{q_i^{1-\alpha} - \lambda(q_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}}}{\sum_{i=1}^n \{q_i^{1-\alpha} - \lambda(q_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}}} \quad (6.1)$$

and λ is to be eliminated using $\sum_{i=1}^n p_i^{\alpha} (q_i^{1-\alpha} - r_i^{1-\alpha}) = 0$.

$$\text{or } \sum_{i=1}^n \{q_i^{1-\alpha} - \lambda(q_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{\alpha}{1-\alpha}} (q_i^{1-\alpha} - r_i^{1-\alpha}) = 0. \quad (6.2)$$

Now we shall first consider two special cases. (i) $\alpha = \frac{1}{2}$
(ii) $\alpha = \frac{2}{3}$.

Case (i) $\alpha = \frac{1}{2}$

Then (6.1) becomes $p_i = \frac{\{\sqrt{q_i} - \lambda(\sqrt{q_i} - \sqrt{r_i})\}^2}{\sum_{i=1}^n \{\sqrt{q_i} - \lambda(\sqrt{q_i} - \sqrt{r_i})\}^2}$ and (6.2) becomes

$$\sum_{i=1}^n \{\sqrt{q_i} - \lambda(\sqrt{q_i} - \sqrt{r_i})\} (\sqrt{q_i} - \sqrt{r_i}) = 0$$

$$\text{or } (1 - \sum \sqrt{r_i q_i}) - 2\lambda(1 - \sum \sqrt{r_i q_i}) = 0 \quad (6.3)$$

$$\text{if } r_i \neq q_i \quad \forall i=1, \dots, n, \quad \lambda = \frac{1}{2}.$$

$$\text{Therefore } p_i = \frac{(\sqrt{q_i} - \sqrt{r_i})^2}{\sum_{i=1}^n (\sqrt{q_i} - \sqrt{r_i})^2} \quad (6.4)$$

Case (ii) $\alpha = \frac{2}{3}$.

In this case, (6.2) reduces to the following quadratic equation in λ

$$\lambda^2 \left\{ \sum_{i=1}^n q_i^{2/3} (q_i^{1/3} - r_i^{1/3}) \right\} + \lambda \left\{ 2 \sum_{i=1}^n (q_i^{1/3} r_i^{1/3} - q_i^{2/3}) (q_i^{1/3} - r_i^{1/3}) \right\} + \left\{ \sum_{i=1}^n (r_i^{2/3} - 2q_i^{1/3} r_i^{1/3}) (q_i^{2/3} - r_i^{1/3}) \right\} = 0 \quad (6.5)$$

The solutions of equation (6.5) are given by

$$\lambda = \frac{- \sum_{i=1}^n (q_i^{2/3} r_i^{1/3} - q_i^{2/3}) (q_i^{1/3} - r_i^{1/3}) \pm \sqrt{\left\{ \sum_{i=1}^n (q_i^{2/3} r_i^{1/3} - q_i^{2/3}) (q_i^{1/3} - r_i^{1/3}) \right\}^2 - \sum_{i=1}^n q_i^{2/3} (q_i^{1/3} - r_i^{1/3}) \left\{ \sum_{i=1}^n (r_i^{2/3} - 2q_i^{1/3} r_i^{1/3}) (q_i^{2/3} - r_i^{1/3}) \right\}}}{\sum_{i=1}^n q_i^{2/3} (q_i^{1/3} - r_i^{1/3})} \quad (6.6)$$

We shall evaluate this case for a set of distributions Q and R and find P .

Example 6.1.1. For $\alpha = \frac{2}{3}$ if we take $Q = (.7, .3)$ and $R = (.4, .6)$, we get the two solutions for (6.6) as $\lambda_1 = 1.5881$ and $\lambda_2 = -0.6575$, and $P_1 = (.2433, .7567)$, $P_2 = (.8491, .1509)$. The corresponding distances are .29 and .42 respectively. So the optimal solution is given by P_1 .

6.2 Problem 2. Out of all distributions that are at a distance a from R , find that distribution which is closest to Q .

Minimize $\frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i q_i^{1-\alpha} - 1 \right\}$ subject to (i) $\sum_{i=1}^n p_i = 1$

$$(ii) \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n r_i^{1-\alpha} - 1 \right\} = a \quad (6.7)$$

By using Lagrange's multipliers method we obtain the minimizing distribution $\{p_i\}$ as follows:

$$p_i = \frac{(q_i^{1-\alpha} - \lambda r_i^{1-\alpha})^{\frac{1}{1-\alpha}}}{\sum_{i=1}^n (q_i^{1-\alpha} - \lambda r_i^{1-\alpha})^{\frac{1}{1-\alpha}}} \quad (6.8)$$

where λ is to be eliminated using (ii) above. The equation for obtaining λ is as follows:

$$\frac{1}{(\alpha-1)} \left\{ \sum_{i=1}^n (q_i^{1-\alpha} - \lambda r_i^{1-\alpha})^{\frac{1}{1-\alpha}} \right\}^{-\alpha} \left\{ \sum_{i=1}^n (q_i^{1-\alpha} - \lambda r_i^{1-\alpha})^{\frac{\alpha}{1-\alpha}} r_i^{1-\alpha} - 1 \right\} = 0 \quad (6.9)$$

Now we shall consider two special cases.

Case (i) $\alpha = \frac{1}{2}$

$$(6.8) \text{ becomes } p_i = \frac{(v q_i - \lambda v r_i)^2}{\sum_{i=1}^n (v q_i - \lambda v r_i)^2} \text{ and } (6.9) \text{ becomes}$$

$$+4 \left\{ \sum_{i=1}^n (v q_i - \lambda v r_i) v r_i - 1 \right\}^2 = a^2 \left\{ \sum_{i=1}^n (v q_i - \lambda v r_i)^2 \right\} \quad (6.10)$$

which is equivalent to

$$(1-a^2)\lambda^2 - 2\lambda \sum_{i=1}^n \sqrt{r_i q_i} (a^2-1) + \left(\sum_{i=1}^n \sqrt{r_i q_i} - 1 \right)^2 - a^2 = 0$$

$$\text{So } \lambda = \frac{+(a^2-1) \sum_{i=1}^n \sqrt{r_i q_i} \pm \sqrt{(a^2-1)^2 \left(\sum_{i=1}^n \sqrt{r_i q_i} \right)^2 + (a^2-1) \left\{ \left(\sum_{i=1}^n \sqrt{r_i q_i} - 1 \right)^2 - a^2 \right\}}}{(1-a^2)}$$

One value of λ here gives the optimal distribution $\{p_i\}$.

Example 6.2.1. $\alpha = \frac{1}{2}$, $Q = (.7, .3)$ and $R = (.4, .6)$. We get the two solutions for (6.11) as $\lambda_1 = .66$, $\lambda_2 = -2.57$ and $P_1 = (.994, .005)$, $P_2 = (.696, .303)$ respectively.

The corresponding distances $D^{1/2}(P_1; Q)$ and $D^{1/2}(P_2; Q)$ are 0.927×10^{-1} and 0.183×10^{-4} respectively. So the optimal solution is given by P_2 .

Case (ii) $\alpha = \frac{2}{3}$.

Then we have from (6.9)

$$-3 \left\{ \sum_{i=1}^n (q_i^{1/3} - \lambda r_i^{1/3})^2 r_i^{1/3} - 1 \right\} = a \left\{ \sum_{i=1}^n (q_i^{1/3} - \lambda r_i^{1/3})^3 \right\}^{2/3}$$

$$\text{or } -27 \left\{ \sum_{i=1}^n (q_i^{1/3} - \lambda r_i^{1/3})^2 r_i^{1/3} - 1 \right\}^3 = a^3 \left\{ \sum_{i=1}^n (q_i^{1/3} - \lambda r_i^{1/3})^3 \right\}^2$$

or with the notation that $\sum_{i=1}^n q_i^{2/3} r_i^{1/3} = A$ and $\sum_{i=1}^n q_i^{1/3} r_i^{2/3} = B$ we have

$$\begin{aligned} & -27[\lambda^6 + 3\lambda^4(A-B-1) + 3\lambda^2\{(A+B)^2 - 2(A+B)\} + \{(A-B)^3 - 3(A+B)^2 + 3(A-B) - 1\}] \\ & = a^3[\lambda^6 - \lambda^5(6B) + 3\lambda^4(3B-2A) - 2\lambda^3(1-9AB) + 3\lambda^2(3A^2+2B) + \lambda(6A) + 1] \end{aligned}$$

or by collecting the coefficients of like powers of λ together, we get

$$\begin{aligned}
& \lambda^6(a^3+27) - \lambda^5(6Ba^3) + 3\lambda^4\{a^3(3B-2A) + 27(A-B-1)\} - 2\lambda^3(1-9AB) \\
& + 3\lambda^2\{3A^2+2B+27(A+B)^2-54(A+B)\} + \lambda(6A)+1+27\{(A-B)^3-3(A+B)^2 \\
& + 3(A-B) - 1\} = 0.
\end{aligned} \tag{6.12}$$

(6.12) is a 6th degree equation in λ one of whose roots gives the required distribution $\{p_i\}$ when substituted in (6.3).

6.3 Problem 3. Out of all probability distributions whose distances from Q and R are in the ratio 1:k find that distribution which is closest to Q.

Solution

Minimize

$$\frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} - 1 \right\}$$

Subject to,

$$(i) \quad \sum_{i=1}^n p_i = 1 \quad \text{and}$$

$$(ii) \quad k \left\{ \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha} - 1 \right\} - \left\{ \sum_{i=1}^n p_i^\alpha r_i^{1-\alpha} - 1 \right\} = 0 \tag{6.13}$$

By making use of Lagrange's multipliers method we obtain the minimizing distribution $\{p_i\}$ as is given below:

$$p_i = \left\{ \frac{[q_i^{1-\alpha} - \lambda(kq_i^{1-\alpha} - r_i^{1-\alpha})]^{\frac{1}{1-\alpha}}}{\sum_{i=1}^n [q_i^{1-\alpha} - \lambda(kq_i^{1-\alpha} - r_i^{1-\alpha})]^{\frac{1}{1-\alpha}}} \right\} \tag{6.14}$$

Now we eliminate λ from the above expression by substituting it in (6.13) above. We obtain the equation for solving λ is follows:

$$\sum_{i=1}^n \frac{\{q_i^{1-\alpha} - \lambda(kq_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{\alpha}{1-\alpha}}}{\left[\sum_{i=1}^n \{q_i^{1-\alpha} - \lambda(kq_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}} \right]^{\alpha}} \{kq_i^{1-\alpha} - r_i^{1-\alpha}\} + (1-k) = 0 \quad (6.15)$$

Case (i) $\alpha = \frac{1}{2}$

Then (6.15) becomes

$$\sum_{i=1}^n \{q_i - \lambda(k\sqrt{q_i} - \sqrt{r_i})\} \{k\sqrt{q_i} - \sqrt{r_i}\} = (k-1) \left[\sum_{i=1}^n \{q_i - \lambda(k\sqrt{q_i} - \sqrt{r_i})\}^2 \right]^{1/2}$$

$$\text{or } \left[\sum_{i=1}^n \{q_i - \lambda(k\sqrt{q_i} - \sqrt{r_i})\} \{k\sqrt{q_i} - \sqrt{r_i}\} \right]^2 = (k-1)^2 \sum_{i=1}^n \{q_i - \lambda(k\sqrt{q_i} - \sqrt{r_i})\}^2$$

or if we denote by $\sigma = \frac{1}{n} \sum_{i=1}^n \sqrt{r_i q_i}$ we get the quadratic equation

$$\begin{aligned} & \lambda^2 \{(k+1)^2 + 4k\sigma(\sigma k - k - 1) - 2(1-k\sigma)(k-1)^2\} \\ & + 2\lambda \{k(3k+1) - \sigma(4k\sigma + k + 1) + (k-1)^2(k+\sigma)\} \\ & + \{k^2 - \sigma(\sigma - 2k) - (k-1)^2\} = 0 \end{aligned} \quad (6.16)$$

and it's solution is given by

$$\lambda = \frac{-I \pm II}{III}$$

$$\text{where } I = -2 \{k(3k+1) - \sigma(4k\sigma + k + 1) + (k-1)^2(k+\sigma)\}$$

$$II = \sqrt{\{I\}^2 - 4 \quad III \quad \{k^2 + \sigma(\sigma - 2k) - (k-1)^2\}}$$

$$\text{and } III = \{(k+1)^2 + 4k\sigma(k - k - 1) - 2(1-k\sigma)(k-1)^2\}.$$

Example 6.3.1. For $\alpha = \frac{1}{2}$, $Q = (.7, .3)$ and $R = (.4, .6)$

we get the two solutions for (6.16) as $\lambda_1 = 11.007$, $\lambda_2 = 0.013$ and $P_1 = (.9268, .0732)$, $P_2 = (.696, .304)$ respectively.

The corresponding distances are $D^{1/2}(P_1:Q)$ and $D^{1/2}(P_2:R)$ are 0.0927 and 0.183×10^{-4} . So the optimal solution is given by P_2 .

Case (ii) $\alpha = \frac{2}{3}$

Equation (6.15) becomes with the notation that

$$A_i = (kq_i^{1/3} - r_i^{1/3})$$

$$\begin{aligned} & \lambda^6 \{ (\Sigma A_i^3)^3 - (1-k)^3 (\Sigma A_i^3)^2 \} - 6\lambda^5 (\Sigma A_i^3) (\Sigma q_i^{1/3} A_i^2) \{ (1-k)^3 (\Sigma A_i^3) - 1 \} \\ & + 3\lambda^4 \{ (\Sigma q_i^{2/3} A_i) (\Sigma A_i^3)^2 + 4(\Sigma A_i^3) (\Sigma A_i^2 q_i^{1/3})^2 - 3(1-k)^3 [(\Sigma q_i^{1/3} A_i^2)^2 \\ & + 2(\Sigma A_i^3) (\Sigma q_i^{2/3} A_i)] \} + 2\lambda^3 \{ (\Sigma q_i^{2/3}) (\Sigma A_i^{1/3} (\Sigma A_i^2 q_i^{1/3}) - \{ (\Sigma A_i^2 q_i^{1/3})^3 \\ & - (1-k)^3 [(\Sigma q_i^{1/3})^3 \Sigma A_i^3 - 9(\Sigma q_i^{1/3} A_i^2) (\Sigma q_i^{2/3} A_i)] \} + 3\lambda^2 \{ (\Sigma q_i^{2/3})^2 (\Sigma A_i^3) \\ & + 4(\Sigma q_i^{2/3} A_i) (\Sigma A_i^2 q_i^{1/3}) - (1-k)^3 [(\Sigma q_i^{2/3} A_i)^2 + (\Sigma q_i^{1/3})^3 (\Sigma q_i^{1/3} A_i^2) \} \\ & - 6\lambda \{ (\Sigma q_i^{2/3} A_i)^2 (\Sigma A_i^2 q_i^{1/3}) + (1-k)^3 (\Sigma q_i^{1/3})^3 (\Sigma q_i^{2/3} A_i) \} + \lambda^0 [\Sigma q_i^{2/3} A_i \\ & - (1-k)^3 (\Sigma q_i^{1/3})^6 \} = 0. \end{aligned}$$

One of the six roots of the above sixth degree equation in λ gives the optimal distribution $\{p_i\}$ when substituted in (6.14).

6.4 Problem 4. Out of all distributions whose sum of distances from Q and R is b , find that distribution which is closest Q .

Solution

$$\text{Minimize } \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i q_i^{1-\alpha} - 1 \right\}$$

$$\text{Subject to (i) } \sum_{i=1}^n p_i = 1$$

(6.17)

$$(ii) \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i^{\alpha} (q_i^{1-\alpha} + r_i^{1-\alpha}) \right\} = b'$$

$$\text{where } b' = b + \frac{2}{\alpha-1}$$

By applying Lagrange's multipliers method we obtain the minimizing distribution $\{p_i\}$ as given below

$$p_i = \frac{\{q_i^{1-\alpha} - \lambda(q_i^{1-\alpha} + r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}}}{\sum_{i=1}^n \{q_i^{1-\alpha} - \lambda(q_i^{1-\alpha} + r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}}} \quad (6.18)$$

and the equation to eliminate λ from (6.8) is

$$\begin{aligned} & \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n \{q_i^{1-\alpha} - \lambda(q_i^{1-\alpha} + r_i^{1-\alpha})\}^{\frac{\alpha}{1-\alpha}} (q_i^{1-\alpha} + r_i^{1-\alpha}) \right\} \\ &= b' \left[\sum_{i=1}^n \{q_i^{1-\alpha} - \lambda(q_i^{1-\alpha} + r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}} \right]^{\alpha} \end{aligned} \quad (6.19)$$

Now we consider the special cases.

$$\text{Case (i) } \alpha = \frac{1}{2}$$

Then we have from (6.19)

$$-2 \left\{ \sum_{i=1}^n (\sqrt{q_i} - \lambda(\sqrt{q_i} + \sqrt{r_i})) (\sqrt{q_i} + \sqrt{r_i}) \right\} = b' \left[\sum_{i=1}^n \{\sqrt{q_i} - \lambda(\sqrt{q_i} + \sqrt{r_i})\}^2 \right]^{1/2}$$

Squaring both sides we get

$$4 \left\{ \sum_{i=1}^n (\sqrt{q_i} - \lambda(\sqrt{q_i} + \sqrt{r_i})) (\sqrt{r_i} + \sqrt{q_i}) \right\}^2 = b'^2 \sum_{i=1}^n \{ \sqrt{q_i} - \lambda(\sqrt{q_i} + \sqrt{r_i}) \}^2$$

which is equivalent to the following quadratic equation in λ ,

with $A_i = (\sqrt{q_i} - \sqrt{r_i})$

$$\lambda^2 (\sum A_i^2) (4 \sum A_i^2 - b'^2) - 2\lambda (4 \sum A_i^2 (1 + \sum \sqrt{r_i} \sqrt{q_i}) - b'^2 \sum A_i \sqrt{q_i}) + (4(1 + \sum \sqrt{r_i} \sqrt{q_i})^2 - b'^2) = 0.$$

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \text{ where } a = (\sum A_i^2) (4 \sum A_i^2 - b'^2)$$

$$b = -2 \{ 4 \sum A_i^2 (1 + \sum \sqrt{r_i} \sqrt{q_i}) - b'^2 \sum A_i \sqrt{q_i} \} \text{ and}$$

$$c = (4(1 + \sum \sqrt{r_i} \sqrt{q_i})^2 - b'^2)$$

Example 6.4.1. For $\alpha = 1/2$, $Q = (.7, .3)$ and $R = (.4, .6)$

we get the two solutions for (6.20) as $\lambda_1 = 0.0457$ and

$\lambda_2 = 11.06$ and $P_1 = (0.713, 0.287)$, $P_2 = (0.545, 0.456)$ respectively.

The corresponding distances $D^{1/2}(P_1:Q)$ and $D^{1/2}(P_2:Q)$ are 0.229×10^{-3} and 0.257×10^{-1} . So the optimal solution is given by P_1 .

Case (ii) $\alpha = \frac{2}{3}$

Then (6.18) becomes

$$-3 \{ \sum q_i^{1/3} - \lambda(q_i^{1/3} + r_i^{1/3}) \}^2 (q_i^{1/3} + r_i^{1/3}) \} = b' [\sum \{ q_i^{1/3} - \lambda(q_i^{1/3} + r_i^{1/3}) \}^3]^{\frac{2}{3}}$$

$$\text{or } -27 \{ \sum \{ q_i^{1/3} - \lambda(q_i^{1/3} + r_i^{1/3}) \}^2 (q_i^{1/3} + r_i^{1/3}) \}^3$$

$$= b'^2 [\sum \{ q_i^{1/3} - \lambda(q_i^{1/3} + r_i^{1/3}) \}^3]^2$$

which is equivalent to the following quartic equation in λ

with the notations

$$B_i = (q_i^{1/3} + r_i^{1/3}) \text{ and } b_1 = (-\frac{b'}{3})^2:$$

$$\begin{aligned} & \lambda^6 \{ (\Sigma B_i^3)^3 - b_1^3 (\Sigma B_i^3)^2 \} \\ & - 6\lambda^5 (\Sigma B_i^3) (\Sigma q_i^{1/3} B_i^2) \{ b_1^3 (\Sigma B_i^3) - 1 \} \\ & + 3\lambda^4 \{ (\Sigma q_i^{2/3} B_i) (\Sigma B_i^3)^2 + 4 (\Sigma B_i^3) (\Sigma B_i^2 q_i^{1/3})^2 \\ & - 3b_1^3 [(\Sigma B_i q_i^{1/3})^2 + 2 (\Sigma B_i^3) (\Sigma B_i q_i^{2/3})] \} \\ & + 2\lambda^3 \{ (\Sigma q_i^{2/3}) (\Sigma B_i^3) (\Sigma B_i^2 q_i^{1/3}) - (\Sigma B_i^2 q_i^{1/3})^3 \\ & - b_1^3 [(\Sigma q_i^{1/3})^3 \Sigma B_i^3 - 9 (\Sigma q_i^{1/3} B_i^2) (\Sigma q_i^{2/3} B_i)] \} \\ & + 3\lambda^2 \{ (\Sigma q_i^{2/3})^2 (\Sigma B_i^3) + 4 (\Sigma q_i^{2/3}) (\Sigma B_i^2 q_i^{1/3}) \\ & - b_1^3 [(\Sigma q_i^{2/3} B_i)^2 + (\Sigma q_i^{1/3})^3 (\Sigma q_i^{1/3} B_i^2)] \} \\ & - 6\lambda \{ (\Sigma q_i^{2/3} B_i)^2 (\Sigma B_i^2 q_i^{1/3}) + b_1^3 (\Sigma q_i^{1/3})^3 (\Sigma q_i^{2/3} B_i) \} \\ & + \lambda^0 \{ \Sigma q_i^{2/3} B_i - b_1^3 (\Sigma q_i^{1/3})^6 \} = 0 \end{aligned} \quad (6.22)$$

One of the six roots of the equation (6.21) gives the optimal probability distribution p_i when substituted in (6.17).

6.5 Problem 5: Find the distribution P from which the sum of distances to be given distributions Q_1, \dots, Q_k

Statement of the problem is minimum.

Solution Minimize

$$\frac{1}{(\alpha-1)} \sum_{j=1}^n \sum_{i=1}^n (p_i q_{ij}^{1-\alpha} - 1)$$

subject to

$$\sum_{i=1}^n p_i = 1 \quad (6.23)$$

Set the Lagrangian

$$L \equiv \frac{1}{\alpha-1} \left(\sum_{j=1}^k \sum_{i=1}^n p_i^{\alpha} q_{ij}^{1-\alpha} - 1 \right) - \lambda \left(\sum_{i=1}^n p_i - 1 \right)$$

Equating $\frac{\partial L}{\partial p_i}$ to zero and solving for p_i using (6.22), we get

$$p_i = \frac{\lambda}{\left[\sum_{j=1}^k q_{ij}^{1-\alpha} \right]^{\frac{1}{\alpha-1}}} \quad (6.24)$$

Substitution of (6.23) in (6.22) results in

$$p_i = \frac{\left[\sum_{j=1}^k q_{ij}^{1-\alpha} \right]^{\frac{1}{1-\alpha}}}{\left[\sum_{i=1}^n \left\{ \sum_{j=1}^k q_{ij}^{1-\alpha} \right\} \right]^{\frac{1}{1-\alpha}}} \quad (6.25)$$

The measure of distance being a convex function of both P and Q (5.24) yields the required optimal probability distribution. We now test it for two special cases.

Case (i) $\alpha = \frac{1}{2}$. Then we have $p_i = \frac{\left(\sum_{j=1}^k \sqrt{q_{ij}} \right)^2}{\sum_{i=1}^n \left\{ \sum_{j=1}^k \sqrt{q_{ij}} \right\}^2}$.

Let us further take $n=2$, $k=3$

$$Q_1 = (.2, .8), Q_2 = (.3, .7) \text{ and } Q_3 = (.4, .6)$$

then we have

$$p_1 = \frac{(\sqrt{.2} + \sqrt{.3} + \sqrt{.4})^2}{(\sqrt{.2} + \sqrt{.3} + \sqrt{.4})^2 + (\sqrt{.6} + \sqrt{.7} + \sqrt{.8})^2} = \frac{(1.543)^2}{(1.543)^2 + (2.51)^2} = .27$$

$$P_2 = \frac{(\sqrt{.6} + \sqrt{.7} + \sqrt{.8})^2}{(\sqrt{.2} + \sqrt{.3} + \sqrt{.4})^2 + (\sqrt{.6} + \sqrt{.7} + \sqrt{.8})^2} = \frac{(2.510)^2}{(1.543)^2 + (2.57)^2}$$

$$= .73$$

The required distribution P is given by (.27, .73).

We shall check it up against an arbitrarily chosen probability distribution, say

$$P_1 = (0.5, 0.5)$$

$$\sum_{j=1}^3 D^{1/2}(P; Q_j) = \frac{1}{\alpha-1} \sum_{j=1}^3 \sum_{i=1}^2 (p_i^{\alpha} q_{ij}^{1-\alpha} - 1) = -2 \{ -3 + \sqrt{.27} \sqrt{.2} + \sqrt{.73} \sqrt{.8}$$

$$+ \sqrt{.27} \sqrt{.3} + \sqrt{.73} \sqrt{.7} + \sqrt{.27} \sqrt{.4} + \sqrt{.73} \sqrt{.6} \}$$

$$= -2 \times -0.014 = 0.028$$

$$\sum_{j=1}^3 D^{1/2}(P_1; Q_j) = -2 \{ -3 + \sqrt{.5} \sqrt{.2} + \sqrt{.5} \sqrt{.8} + \sqrt{.5} \sqrt{.3} + \sqrt{.5} \sqrt{.7} + \sqrt{.5} \sqrt{.4} + \sqrt{.5} \sqrt{.6} \}$$

$$= -2 \{ -3 + \sqrt{.5} (\sqrt{.2} + \sqrt{.3} + \sqrt{.3} + \sqrt{.7} + \sqrt{.4} + \sqrt{.6}) \}$$

$$= -2 \times (-0.077) = 0.154$$

$$\text{Hence we have } \sum_{j=1}^3 D^{1/2}(P_1; Q_j) > \sum_{j=1}^3 D^{1/2}(P; Q_j).$$

Case (ii) $\alpha = \frac{2}{3}$

$$\text{Then we have } p_i = \frac{\left\{ \sum_{j=1}^3 (q_{ij})^{1/3} \right\}^3}{\sum_{i=1}^2 \left\{ \sum_{j=1}^3 q_{ij}^{1/3} \right\}^3}.$$

We shall evaluate p_i taking the same distributions Q_1, Q_2 and Q_3 as in case (i).

$$p_1 = \frac{((.2)^{1/3} + (.3)^{1/3} + (.4)^{1/3})^3}{((.2)^{1/3} + (.3)^{1/3} + (.4)^{1/3})^3 + (.6)^{1/3} + (.7)^{1/3} + (.8)^{1/3})^3}$$

$$= \frac{7.9896}{7.9896 + 18.8814} = 0.2973317$$

$$p_2 = \frac{(.6)^{1/3} + (.7)^{1/3} + (.8)^{1/3}}{(.2)^{1/3} + (.3)^{1/3} + (.4)^{1/3} + (.6)^{1/3} + (.7)^{1/3} + (.8)^{1/3}}$$

$$= \frac{18.8814}{7.9896 + 18.8814} = 0.7026683$$

So the required distribution is $P = (0.29, 0.71)$.

$$\begin{aligned} \text{Now } \sum_{j=1}^3 D^{1/3}(P; Q_j) &= -\frac{3}{2} \{ -3 + (.29)^{2/3} (.2)^{1/3} + (.71)^{2/3} (.8)^{1/3} \\ &\quad + (.29)^{2/3} (.3)^{1/3} + (.71)^{2/3} (.7)^{1/3} \\ &\quad + (.29)^{2/3} (.4)^{1/3} + (.71)^{2/3} (.6)^{1/3} \} \\ &= -\frac{3}{2} \{ -3 + (.29)^{2/3} (.2^{1/3} + .3^{1/3} + .4^{1/3}) \\ &\quad + (.71)^{2/3} (.6^{1/3} + .7^{1/3} + .8^{1/3}) \} \\ &= -1.5 \{ -3 + 2.98 \} = 0.0293 \end{aligned}$$

$$\begin{aligned} \text{and } \sum_{j=1}^3 D^{1/3}(P_1; Q_j) &= -1.5 \{ -3 + (0.5)^{2/3} (1.98533 + 2.63689) \} \\ &= 0.112 \end{aligned}$$

$$\text{Again } \sum_{j=1}^3 D^{1/3}(P_1; Q_j) > \sum_{j=1}^3 D^{1/3}(P; Q_j).$$

6.6 Problem: 6 Out of all distributions from which the sum of weighted distances to Q_1, \dots, Q_m is a constant find the distribution that is closest to Q_0 .

Solution Minimize

$$\frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i q_{i0}^{\alpha-1} - 1 \right\}$$

subject to (i) $\sum_{i=1}^n p_i = 1$

(6.26)

$$(ii) \frac{1}{\alpha-1} \sum_{j=1}^m \lambda_j \left(\sum_{i=1}^n p_i q_{ij}^{\alpha-1} - 1 \right) = T$$

By applying Lagrange's multipliers method, we get the minimizing distribution $\{p_i\}$ as given below:

$$p_i = \frac{(q_{i0}^{\alpha-1} - A \sum_{k=1}^m \lambda_k q_{ik}^{\alpha-1})^{\frac{1}{1-\alpha}}}{\sum_{i=1}^n (q_{i0}^{\alpha-1} - A \sum_{k=1}^m \lambda_k q_{ik}^{\alpha-1})^{\frac{1}{1-\alpha}}}$$

(6.27)

Now A is to be determined from the following equation:

$$\frac{1}{\alpha-1} \sum_{j=1}^m \lambda_j \left(\sum_{i=1}^n \frac{(q_{i0}^{\alpha-1} - A \sum_{k=1}^m \lambda_k q_{ik}^{\alpha-1})^{\frac{\alpha}{1-\alpha}}}{\{ \sum_{i=1}^n (q_{i0}^{\alpha-1} - A \sum_{k=1}^m \lambda_k q_{ik}^{\alpha-1})^{\frac{1}{1-\alpha}} \}} q_{ij}^{\alpha-1} - 1 \right) = T$$

$$\text{or } \frac{1}{\alpha-1} \left[\sum_{j=1}^m \lambda_j \left(\sum_{i=1}^n \{ q_{i0}^{\alpha-1} - A \sum_{k=1}^m q_{ik}^{\alpha-1} \}^{\frac{\alpha}{1-\alpha}} q_{ij}^{\alpha-1} \right) \right] = T', \quad \left\{ \sum_{i=1}^n (q_{i0}^{\alpha-1} - A \sum_{j=1}^m q_{ij}^{\alpha-1})^{\frac{1}{1-\alpha}} \right\}^{\alpha}$$

(6.28)

where $T' = T + \frac{1}{\alpha-1}$.

Now we shall consider the special case: $\alpha = \frac{1}{2}$

Then (6.27) becomes

$$-2 \left[\sum_{j=1}^m \lambda_j \left(\sum_{i=1}^n \{ \sqrt{q_{i0}} - A \sum_{j=1}^m \sqrt{q_{ij}} \} q_{ij} \right) \right] = T' \left\{ \sum_{i=1}^n \left(\sqrt{q_{i0}} - A \sum_{j=1}^m \sqrt{q_{ij}} \right)^2 \right\}^{\frac{1}{2}} \quad (6.29)$$

$$\text{or } 4 \left[\sum_{j=1}^m \lambda_j \left(\sum_{i=1}^n \left(\sqrt{q_{i0}} - A \sum_{j=1}^m \sqrt{q_{ij}} \right) \right) \right]^2 = T' \left\{ \sum_{i=1}^n \left(\sqrt{q_{i0}} - A \sum_{j=1}^m \sqrt{q_{ij}} \right)^2 \right\}$$

$$A^2 \left\{ 4 \left(\sum_{j=1}^m \sum_{i=1}^n \sum_{k=1}^m \lambda_j \sqrt{q_{ik} q_{ij}} \right)^2 - T' \left\{ \sum_{i=1}^n \sum_{j=1}^m \sqrt{q_{i0} q_{ij}} \right\}^2 \right\}$$

$$- 2A \left\{ \left(\sum_{j=1}^m \sum_{i=1}^n \lambda_j \sqrt{q_{i0} q_{ij}} \right) \left(\sum_{j=1}^m \sum_{i=1}^n \sum_{k=1}^m \lambda_j \sqrt{q_{ik} q_{ij}} \right) \times 4 \right.$$

$$\left. T' \left\{ \left(\sum_{i=1}^n \sqrt{q_{i0}} \right) \left(\sum_{i=1}^n \sum_{j=1}^m \sqrt{q_{ij} q_{i0}} \right) \right\} + \left\{ 4 \left(\sum_{j=1}^m \sum_{i=1}^n \sqrt{q_{i0} q_{ij}} \right) \right\} \right.$$

$$\left. - T' \left\{ \left(\sum_{i=1}^n \sqrt{q_{i0}} \right)^2 \right\} \right\} = 0 \quad (6.30)$$

(6.29) gives two values of A one of which will give as the required optimal distribution p_i when substituted in (6.26)

$$A = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

$$\text{where } a = 4 \left(\sum_{j=1}^m \sum_{i=1}^n \sum_{k=1}^m \lambda_j \sqrt{q_{ik} q_{ij}} \right)^2 - T' \left\{ \left(\sum_{i=1}^n \sum_{j=1}^m \sqrt{q_{i0} q_{ij}} \right)^2 \right\}$$

$$b = -2 \left\{ \left(\sum_{j=1}^m \sum_{i=1}^n \sqrt{q_{i0} q_{ij}} \lambda_j \right) \left(\sum_{j=1}^m \sum_{i=1}^n \sum_{k=1}^m \lambda_j \sqrt{q_{ik} q_{ij}} \right) 4 \right.$$

$$\left. - T' \left\{ \left(\sum_{i=1}^n \sqrt{q_{i0}} \right) \left(\sum_{i=1}^n \sum_{j=1}^m \sqrt{q_{ij} q_{i0}} \right) \right\} \right\}$$

$$\text{and } c = \left\{ 4 \left(\sum_{i=1}^n \sum_{j=1}^m \lambda_j \sqrt{q_{i0} q_{ij}} \right)^2 - T' \left\{ \left(\sum_{i=1}^n \sqrt{q_{i0}} \right)^2 \right\} \right\}.$$

6.7 Problem 7. Out of all distributions which are equally close to Q and R find that distn. which is closest to another distribution S.

Solution: Minimize $\frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i^\alpha s_i^{1-\alpha} - 1 \right\}$ subject to

$$(i) \quad \sum_{i=1}^n p_i = 1$$

$$(ii) \quad \frac{1}{\alpha-1} \left\{ \sum_{i=1}^n p_i^\alpha (q_i^{1-\alpha} - r_i^{1-\alpha}) \right\} = 0 \quad (6.31)$$

By applying Lagrange's method, we get the optimal distribution $\{p_i\}$ is given by

$$p_i = \frac{\{s_i^{1-\alpha} - \lambda(q_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}}}{\sum_{i=1}^n \{s_i^{1-\alpha} - \lambda(q_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}}} \quad (6.32)$$

Now λ is to be eliminated by using (6.31) and (6.30): we get after the substitution

$$\left\{ \sum_{i=1}^n \frac{\{s_i^{1-\alpha} - \lambda(q_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{\alpha}{1-\alpha}}}{\left[\sum_{i=1}^n \{s_i^{1-\alpha} - \lambda(q_i^{1-\alpha} - r_i^{1-\alpha})\}^{\frac{1}{1-\alpha}} \right]^\alpha} (q_i^{1-\alpha} - r_i^{1-\alpha}) \right\} = 0 \quad (6.33)$$

Case (i) Let $\alpha = \frac{1}{2}$. Then (6.32) becomes

$$\sum_{i=1}^n \{ \sqrt{s_i} - \lambda(\sqrt{q_i} - \sqrt{r_i}) \} (\sqrt{q_i} - \sqrt{r_i}) = 0$$

$$\text{or } \sum \sqrt{q_i} s_i - \sum \sqrt{r_i} s_i - 2\lambda(1 + \sum \sqrt{r_i} q_i) = 0 \quad (6.34)$$

From (6.33), $\lambda = \frac{\sum \sqrt{s_i} (\sqrt{q_i} - \sqrt{r_i})}{2(1 + \sum \sqrt{r_i q_i})}$ which when substituted in

(6.31) gives the required minimizing distribution $\{p_i\}$.

Case (ii) $\alpha = \frac{2}{3}$. Then we have from (6.32)

$$\sum_{i=1}^n \{s_i^{1/3} - \lambda(q_i^{1/3} - r_i^{1/3})\}^2 (q_i^{1/3} - r_i^{1/3}) = 0$$

Let $A_i = (q_i^{1/3} - r_i^{1/3})$, then the above equation becomes

$$\sum_{i=1}^n \{s_i^{2/3} + \lambda^2 A_i^2 - 2\lambda s_i^{1/3} A_i\} = 0$$

or the quadratic equation in λ :

$$\lambda^2 \left(\sum_{i=1}^n A_i^2 \right) - 2\lambda \left(\sum_{i=1}^n s_i^{1/3} A_i \right) + \left(\sum_{i=1}^n s_i^{2/3} A_i \right) = 0$$

whose solutions are given by

$$\lambda = \frac{\sum_{i=1}^n s_i^{1/3} A_i^2 \pm \sqrt{\left(\sum_{i=1}^n s_i^{1/3} A_i^2 \right)^2 - \left(\sum_{i=1}^n A_i^2 \right) \left(\sum_{i=1}^n s_i^{2/3} A_i \right)}}{\left(\sum_{i=1}^n A_i^2 \right)}$$

One of these roots will give the optimal distribution $\{p_i\}$ when substituted in (6.31).

With that we come to the end of this chapter.

References

1. J. Aczel and Z. Daroczy, "Charakterisierung de Entropism Positive Ordiung und der Shannon Schen Entropis", Acta. Math. Acad. Sci., Hungary, Vol. 14(1963), pp. 95-121.
2. P.N. Arora, "An Axiomatic Characterization of Inaccuracy Functions", Ind. Journ. Pure Appl. Math., Vol. 13(1982), pp. 158-162.
3. M. Behara and J.S. Chawla, "Generalized Gamma Entropy", In Selecta. Statistica Canadiana, Vol. 2(1974), pp 15-38.
4. B. Bessat and A. Raviv, "Renyi's Entropy and Probability of Error", IEEE. Trans.I.T., 24(1978),pp
5. A.B. El-Sayeed, "The Independence Inequality and its Applications to Information Theory" Information and Control, Vol. 35(1977), 229-245.
6. A.B. El-Sayeed, "On Properties of Measures of Information", Journ. of Information and Optimization Sciences, Vol. 2 (1981), pp. 189-118.
7. C. Ferrari, "Hypoentropy and Related Heterogenaty, Divergence and Information Measures", Statistica (Bologria), Vol. 40(1980), pp. 155-168.
8. J. Havrda and F. Charavet, " Quantification Method of Qualification Processes: Concept of Structured X-Entropy" Kybernetika, Vol. 3(1967), pp. 30-35.
9. J.N. Kapur, "Generalized Entropies of Order α and Type β " The Math. Seminar, Vol. 4(1967), pp. 78-94.
10. _____, "Information of Order α and Type β ", Proc. Ind. Acad. Sci. Vol. 69A(1968), pp. 201-211.
11. _____, "Twenty Five Years of Maximum Entropy Principle", Journ. Math. Phy. Sciences, Vol. 17(1983), pp 103-156.
12. _____, "Non-Additive Measures of Entropy and Distributions of Statistical Mechanics", Ind. Journ. Pure and Appl. Math., Vol. 14(1983), pp. 1372-1387.

13. J.N. Kapur, 'A Comparative Assessment of Various Measures of Entropy', Journ. Info. and Optimization Sciences, Vol. 4(1983), pp. 207-232.
14. _____, 'A Comparative Assessment of Various Measures of Directed Divergence', Advances in Management Studies, Vol. 3(1984), pp. 1-16.
15. J.N. Kapur, B.L. Bhatia and C.R. Bector, 'On Entropy and Directed Divergence of Order α and Type β ', Adv. Magt. Studies, Vol.3(1984), pp. 129-144.
16. J.N. Kapur, 'The Roles of Maximum Entropy and Minimum Discrimination Information Principles in Statistics', Journ. Ind. Soc. Ag. Stat., Vol.36(3)(1984), pp. 12-58.
17. _____, 'Four Families of Measures of Entropy', Ind. Journ. Pure and Appl. Math. 17(1986), pp. 429-449.
18. _____, 'On Monotonocity Property of Various Measures of Entropy and Directed Divergence' Journ. of Math. and Phy. Sci. Vol.20(1986), pp. 239-257.
19. _____, 'On the Upper Bounds for Jensen's Difference for Various Measures of Entropy', Nat. Acad. Sci. Letts. Vol.9(1986), pp. 351-354.
20. _____, 'On Convex-Semi Metric Spaces and Generalized Shannon's Inequalities', I.I.T., Kanpur Res. Rep. No. 193.
21. _____, 'Some new Measures of Entropy and Divergence', I.I.T., Kanpur Res. Rep. No. 236.
22. _____, 'A Critique of Certain Measures of Entropy', I.I.T., Kanpur Res. Rep. No. 237.
23. _____, 'Some New-Measures of Inaccuracy', I.I.T., Kanpur Res. Rep. No. 246.
24. _____, 'On the Validity of Certain Generalized Measures of Information, Directed Divergence and Inaccuracy', I.I.T., Kanpur Res. Rep. No. 250.
25. _____, 'On the Relationship Between Measures of Entropy and Directed Divergence', I.I.T., Kanpur Res. Rep. No. 251.

26. J.N. Kapur, "Some New Measures of Directed Divergence", I.I.T., Kanpur Res. Rep. No. 252.
27. _____, "New Measures of Entropy Divergence and Inaccuracy", I.I.T., Kanpur Res. Rep. No. 254.
28. _____, "Measures Information: A Broad Survey", I.I.T., Kanpur Res. Rep. No. 265.
29. _____, "Some Normalized Measures of Directed Divergence", I.I.T., Kanpur Res. Rep. No. 269.
30. _____, "On the Validity of Sharma and Mittal's Measures of Entropy" I.I.T., Kanpur Res. Rep. No. 271.
31. _____, "On the Range of Validity of Sharma and Taneja's Measures of Entropy", I.I.T., Kanpur Res. Rep. No. 272.
32. _____, "On the Range of Validity of Entropy of Order α and Type β ", I.I.T., Kanpur Res. Rep. No. 273.
33. _____, "Some new Non-Additive Measures of Entropy" I.I.T., Kanpur Res. Rep. No. 284.
34. _____, "Monotonocity and Concavity of Some Parametric Measures of Entropy Directed Divergence and Related Functions", I.I.T., Kanpur Res. Rep. No. 301.
35. _____, "Seven Optimization Problems in Information Theory", I.I.T., Kanpur Res. Rep. No. 309.
36. _____, "Two Parametric Measures of Directed Divergence", I.I.T., Kanpur Res. Rep. No. 310.
37. _____, "Some New Three Parametric Measures of Entropy and Directed Divergence", I.I.T., Kanpur Res. Rep. No. 325.
38. _____, "On Measures of Directed Divergence based on Jensen's Difference", I.I.T., Kanpur Res. Rep. No. 330.
39. _____, "A Result in the Geometry of Probability Space", I.I.T., Kanpur Res. Rep. No. 335.
40. _____, "New Generalized Measures of Entropy and Directed Divergence", I.I.T., Kanpur Res. Rep. No. 341.

41. J.N. Kapur, 'Importance of the Concept of Distance in Probability Space', I.I.T., Kanpur Res. Rep. No. 360.
42. _____, 'On Some Useful Properties of Measures of Entropy and Directed Divergence' I.I.T., Kanpur Res. Rep. No. 362.
43. _____, 'Maximum Entropy Models in Science and Engineering' South Asia Publishers (1984), New Delhi.
44. D.F. Kerridge, 'Inaccuracy and Inference' J. Roy. Stat. Soc. Ser. B, Vol. 23(1961), pp. 183-194.
45. S. Kullback and R.A. Leibler, 'On Information and Sufficiency' Ann. Math. Stat. Vol. 22(1951), pp 79-86.
46. S. Kullback, 'Information Theory and Statistics' John Wiley, New York (1959).
47. A.M. Mathai and P.N. Rathie, 'Basic Concepts in Information Theory' Wiley Eastern Ltd., New Delhi (1975).
48. P. Nath, 'Inaccuracy and Coding Theory' Metrika Band. 24(1977), pp. 123-135.
49. _____, 'Entropy, Inaccuracy and Information' Metrika Band 24(1977), pp 136-148.
50. G.C. Patni and K.C. Jain, 'On Axiomatic Characterization of Some non-Additive Measures of Information' Metrika, Band 24(1977), pp 23-34.
51. P.N. Rathie, 'On a Generalized Entropy and a Coding Theorem' Journ. App. Prob., Vol. 7(1970), pp. 124-133.
52. P.N. Rathie, A Generalization of the non-additive Measure of Uncertainty and Information, Kybornetika Vol. 76(1971), pp. 125-132.
53. _____, 'A Generalization of the non-Additive Measures of Uncertainty and Information' Kybernetika, Vol. 7(1971), pp. 125-132.
54. P.N. Rathie and P. Kannappan, 'An Inaccuracy Function of Type β ' Am. Inst. Stat. Math. Vol. 25(1975), pp. 205-214.

55. A. Renyi, "'On Measures of Entropy and Information'"
Proc. 4th Berkeley Symp. Math. Stat. and Prob.,
Vol. 1 (1961), pp. 547-561.
56. C.E. Shannon, "'A Mathematical Theory of Communication'"
Bell. Syst. Tech. Journ. Vol. 27(1948), pp 379-423,
623-659.
57. B.D. Sharma and I.J. Taneja, "'Entropy of Type (α, β)
and other Generalized Measures in Information
Theory'" Metrika Vol. 22(1975), pp. 205-215.
58. B.D. Sharma and D.P. Mittal, "'New Non-Additive Measures
of Entropy for Discrete Probability Distribution'"
Journ. Math. Sci. Vol. 10(1975), pp. 28-40.
59. B.D. Sharma and H.C. Gupta, "'Sub Additive Measures of
Relative Information and Inaccuracy'" Metrika,
Band 23(1976), pp. 155-165.
60. J.C.A van der Lubbe, "'On Certain Coding Theorems for
Information of Order α and Type β '" In, Trans. 8th
Prague Conf. Inf. Theo., D. Reidel, Dordrecht
(1978), pp. 253-266.
61. J.C.A. van der Lubbe, Y. Boxma and D.E. Bockee,
"'A Generalized Class of Certainty and Information
Measures'" Information Sciences, Vol. 32(1984),
pp. 187-215.